Estimating the Gains (and Losses) of Revenue Management*

Xavier D'Haultfœuille[†] Ao Wang[‡] Philippe Février[§]
Lionel Wilner[¶]

June 9, 2022

Abstract

Despite the wide adoption of revenue management in many industries such as airline, railway, and hospitality, there is still scarce empirical evidence on the gains or losses of such strategies compared to uniform pricing or fully flexible strategies. We quantify such gains and losses and identify their underlying sources in the context of French railway transportation. The identification of demand is complicated by censoring and the absence of exogenous price variations. We develop an original identification strategy combining temporal variations in relative prices, consumers' rationality and weak optimality conditions on the firm's pricing strategy. Our results suggest similar or better performance of the actual revenue management compared to optimal uniform pricing, but also substantial losses of up to 16.2% compared to the optimal pricing strategy. We also highlight the key role of revenue management in acquiring information when demand is uncertain.

Keywords: Revenue management, dynamic pricing, demand estimation, demand learning, moment inequalities.

JEL Codes: C25, C61, D12, R40.

*We thank people at iDTGV for providing us the data. We also thank Philippe Choné, Pierre Dubois, Carlos Daniel Santos, Laurent Lamy, Matt Shum and participants of seminars at California Institute of Technology, Mannheim University, Rice University, Sciences-Po, Texas A&M, Toulouse School of Economics, Yale University and EARIE in Athens, 3rd CREST-ECODEC Conference, the 5th French Econometrics Conference for their comments.

[†]CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr.
[‡]University of Warwick and CAGE, ao.wang@warwick.ac.uk.
[§]CREST, fevrier@ensae.fr.
[¶]CREST, lionel.wilner@ensae.fr.

1 Introduction

Revenue management, namely the practice of adjusting supply to the random demand for perishable goods, is an old practice, which has increased in importance with the rise of the e-commerce (Boyd and Bilegan, 2003). Adjusting prices in a flexible way is likely to increase firms' revenues but it also comes at some cost, as it requires specialized teams and good algorithms. This continuous updating is also a complex exercise, so simple rules are usually set to simplify the pricing strategy. These rules may be suboptimal. This paper identifies how much gains in revenues can be expected by firms when adopting flexible strategies compared to uniform pricing. We also quantify the magnitude of losses of actual strategies compared to the optimal ones, under various constraints imposed on such strategies. Finally, by varying these constraints and the assumptions behind the counterfactuals, we identify the main sources of the gains or losses.

We address these questions by studying revenue management at iDTGV, a subsidiary of the French railway monopoly, SNCF. From 2004 to 2017, this firm provided low-cost trains from Paris to several towns in France, and the corresponding return trains. Its revenue management was based on quantities, as is often the case in companies selling perishable goods (e.g. flight tickets, hotel rooms, rented cars for given periods etc.).[1] Namely, for the economy class on which we focus here, 12 classes of prices, called fare classes hereafter, were defined. These 12 prices were sorted in ascending order and for a given trip (e.g. Paris-Bordeaux), set almost constant during the period we studied. For each train, revenue managers could decide, at any moment before the departure and depending on the demand, to close the current fare class and open the next one, thus increasing the prices of the seats. We investigate hereafter the relative benefits of this popular pricing strategy compared to uniform pricing, or alternative, more flexible strategies.

In order to compute such counterfactuals, we first show that in our context, recovering the price elasticity coefficient, relative demand parameters (of, e.g. Bordeaux versus Toulouse in Paris-Toulouse trains) and the total demand for a given train at a given price are sufficient to identify a rich set of counterfactual revenues. In par-

[1] For a detailed review of revenue management techniques, see Talluri and van Ryzin (2005).

ticular, the timing of consumers' arrival is not necessary to identify counterfactual revenues. This is convenient because such information is often unobserved, as in our case. We can compute revenues not only under uniform pricing, but also under optimal dynamic pricing, with any number of fare classes. Importantly also, we are able to compute such counterfactuals assuming either that iDTGV has complete or incomplete information on a given train's demand.

The identification of price elasticity, relative demand parameters and the total demand at a given price is however complicated by two issues that are likely to arise in many markets of perishable goods. First, and as already observed by Swan (1990), Lee (1990) and Stefanescu (2012), we face a severe censoring problem: demand at a given price is generally larger than the number of seats sold at that price. Second, prices vary only within the grids of 12 prices corresponding to each of the 12 fare classes. Hence, we cannot rely on usual instruments such as cost shifters.

To identify price elasticity, we rely on a new argument tailored to our application but that may apply to other contexts as well. Specifically, we exploit the fact that revenue management is done at a route level (e.g. Paris-Toulouse), while the train serves several cities (e.g. Bordeaux and Toulouse). This means that fare classes close at the same time for all destinations within the same route. Relative prices between, e.g. Bordeaux and Toulouse, then vary simultaneously whenever a fare class closes.[2] We prove that the price elasticity can be identified by relating variations between relative prices and the proportion of consumers buying tickets for one destination versus another. Specifically, identification can be achieved under the assumption that price elasticities and the proportion of consumers seeking to buy a ticket for one destination versus another remain constant over time. We can test both conditions empirically, and the results suggest that they are reasonable in our context.

The identification of the distribution, over the different trains, of the total demand at a given price is also difficult, in particular because of the censoring problem mentioned above. We first show that basic conditions on the rationality of consumers deliver inequalities relating this total demand with the number of seats that are sold. We complement these inequalities by weak optimality conditions on the observed revenue

[2] A similar strategy could be used for, e.g., hotels, if the prices of rooms of different qualities change simultaneously.

management. Specifically, we assume that this revenue management was better, on average, than a uniform pricing strategy performed under incomplete information and using prices from the observed grid of prices. Given our very purpose, it is important here not to impose too strong optimality conditions, such as optimality vis-à-vis all dynamic strategies, as these conditions would very much drive our results. Also, our conditions have the advantage of being relatively simple to exploit for identification and estimation purposes. At the end, these conditions stemming from demand and supply can be combined to form a set of moment inequalities. Though they rely on weak restrictions, these moment inequalities are sufficient to produce informative bounds on most counterfactual revenues.

We obtain the following key findings. First, we estimate a price elasticity of about -4, which is below the range of most estimates in the transportation literature (see, e.g. Jevons et al., 2005, for a meta-analysis). However, we show in Appendix B that using aggregated quantities and prices to estimate price elasticity, as done by most of these studies, produces estimates that are substantially biased towards zero. Second, our results suggest that the observed revenue management practice was effective but still sub-optimal. The observed revenue management generated a gain of up to 8.1% compared to the optimal uniform pricing in an incomplete information set-up. However, we also estimate, under the same informational set-up, a loss of at least 6.8% and up to 15.1% compared to the optimal pricing strategy under the same restriction of 12 ascending fare classes as those actually used. Actually, we estimate that simple strategies, such as 12 (non necessarily ascending) fare classes, already secure almost 99% of the fully unconstrained optimal pricing strategy.

Lastly, we emphasize the key role of demand uncertainty on revenues, and how revenue management can mitigate it. Revenues from a uniform pricing strategy are 17.2% higher when moving from an incomplete to a complete information set-up. But the informational gains are much smaller (0.22%) when considering fully flexible pricing strategies. In other words, implementing the optimal dynamic pricing strategy mitigates almost entirely the loss entailed by demand uncertainty. The reason behind is that information accumulates quickly: by observing and learning from the sales of half of the available seats, the firm can already secure more than 97% of the revenue under complete information.

Related Literature. Our paper relates to several theoretical and empirical papers in operational research and economics. The theoretical literature on revenue management has investigated optimal quantity-based revenue managements, where firms segment demand by choosing either once for all or dynamically the allocation of, say, seats into fare classes in which prices are predetermined. We refer in particular to Littlewood (1972) and Brumelle and McGill (1993) for static solutions, and to Gallego and Van Ryzin (1994), Feng and Gallego (1995), Feng and Xiao (2000), Aviv and Pazgal (2002) for dynamic solutions. These last papers have studied optimal pricing strategies assuming that consumers arrive under some homogeneous Poisson process.

In our paper, we assume that consumers arrive according to a flexible, non-homogenous Poisson process, as Bitran and Mondschein (1997), Zhao and Zheng (2000), and McAfee and te Velde (2008). Our demand model is closest to McAfee and te Velde (2008), but with one key difference. Whereas they assume that the firm has a complete information on the demand parameters, we also consider an incomplete information set-up where only the distribution of these parameters is known. The firm then updates this distribution as consumers arrive. Such an incomplete information set-up seems more plausible when, as here, aggregate demand may vary much from one train to another. We also generalize McAfee and te Velde (2008) by studying constrained pricing strategies close to those implemented in practice. We refer to Online Appendix D for details on the resolution of the corresponding Bellman equations.

Our results underline the important role of information and demand learning to explain the gains and losses of revenue management. Such a point has already been made in the theoretical literature but to our knowledge, we are the first to quantify these roles using real data.[3] Lin (2006) studies similar models to ours in his sections 5.1 and 5.2 and allows for firm's Bayesian learning from the observed purchases or arrivals. Instead of deriving the optimal policy, his paper focuses on a specific policy (variable-rate), which is shown to be nearly optimal in simulations. Aviv and Pazgal (2002) derives the optimal policy assuming an unknown constant arrival rate of consumers and simulates the loss due to incomplete information. By contrast, we allow

[3]Different from the revenue management we consider, Huang et al. (2022) study how firms' static pricing in liquor market in the US can be improved by learning market conditions from realized sales.

for heterogeneous arrival rate, and study other practically relevant pricing strategies as well. Finally, in contrast to all these papers and ours, den Boer and Zwart (2015) consider another form of learning by the firm, based on maximizing the likelihood of the data at its disposal. We refer to den Boer (2015) for a complete survey on demand learning in dynamic pricing.

In the empirical literature on revenue management, the closest papers to ours are Lazarev (2013) and Williams (2022), both of which study dynamic airline pricing in a monopolistic market.[4] While both papers accentuate price discrimination and its welfare effects, the main goal of our paper is to quantify the potential gains and loss due to revenue management in practice. As a result, contrary to their models, ours explicitly incorporates firm's learning behavior from the realized demand. Moreover, we do not impose strong optimality conditions on the observed prices.[5] On the other hand, while Lazarev (2013) allows travelers to be forward-looking, we abstract from any strategic considerations from consumers here, following Williams (2022) and the operation research literature. The rationale behind is that in our context, and contrary to what happens in the airline industry, prices always increase. So at least in the absence of uncertainty on the opportunity of the journey, the consumers have no incentives to wait.

The rest of the paper is organized as follows. In Section 2, we present the context and our data. Section 3 displays the demand model and our assumptions on the supply side. Section 4 is devoted to the identification and estimation of demand under our assumptions and given the data at our disposal. Section 5 presents the results. The appendix gathers the proofs of our identification results and estimation with aggregate data. The Online Appendix displays the formulas for the counterfactual revenues, additional details on some robustness checkcs and additional proofs.

[4]Another recent empirical paper is Cho et al. (2018), which studies revenue management under oligopoly in hospitality industry. Their analysis focuses on the pricing behavior of "hotel 0" (from which the demand data is obtained) in a competing environment.

[5]See also Cho et al. (2018, 2019) for recent examples that identify demand without imposing strong optimality conditions.

2 Institutional Background and Data

2.1 Revenue Management at iDTGV in 2007-2009

iDTGV was a low-cost subsidiary of the French railway monopoly, SNCF, which was created in 2004 and disappeared in December 2017.[6] It owned its trains and had a pricing strategy independent from SNCF. Prices were generally lower than the full-rate prices of SNCF, but were also associated with a slightly lower quality of services. Namely, tickets could only be bought on Internet, they were nominative and could not be cancelled. On top of that, they could be exchanged only under some conditions and at some cost.

The routes of iDTGV were all between Paris and other towns. For each of those towns and every day, one train was leaving Paris and another coming to Paris. Table 1 presents the routes we observe in our data from May 2007 till March 2009. These routes have several stops, but to simplify the analysis, we gather them so as to form a single intermediate stop and a single final stop. We aggregate the cities according to the price schedule. For instance, we group Aix-en-Provence and Avignon together in the Paris-Marseille route since the corresponding prices are always the same. This gathering is consistent with Assumption 1 below, as our demand model remains valid after aggregation of cities.

Different routes may share the same intermediate destination. For instance, Bordeaux is the intermediate destination of Paris-Côte basque and Paris-Toulouse. Importantly, no tickets were sold between the intermediate and the final destination, e.g. no Bordeaux-Toulouse tickets are sold on the Paris-Toulouse route. Our understanding is that this was done to avoid controlling people in intermediate destinations, as there were no ticket inspectors in the trains.

[6] Its disappearance was due to internal strategic considerations at SNCF. It was basically replaced by Ouigo, the new low-cost service at SNCF.

Table 1: Routes with intermediate and final destinations

Route name	Final stop(s)	Intermediate stop(s)	Nb. of trains
Côte d'Azur	Cannes,Saint-Raphaël,Nice	Avignon	452
Marseille	Marseille	Aix-en-Provence/Avignon	453
Perpignan	Perpignan	Nîmes, Montpellier	689
Côte basque	St Jean de Luz,Bayonne, Biarritz,Hendaye	Bordeaux	405
Toulouse	Toulouse	Bordeaux	411
Mulhouse	Mulhouse	Strasbourg	499
Total			2,909

Notes: we have different number of observations for the different routes because the period we cover varies slightly from one route to another.

The trains are split into economy class and business class cars of fixed sizes. Revenue management was implemented almost independently between the two classes, i.e. under the sole constraint that prices in economy class are always lower than in business class. This constraint was very seldom binding in practice, so we ignore it hereafter. We focus hereafter on the economy class, which represents roughly 73% of the seats. In this category, there are 12 fare classes corresponding to 12 prices sorted in ascending order. The price of a given fare class, at a peak time or off peak and for some origin-destination trip (e.g. Paris-Bordeaux) remained constant for several months (e.g. from 03/01/2007 to 10/31/2007) before being adjusted marginally, mostly to account for inflation. Contrary to SNCF, iDTGV did not make any third-degree price discrimination, so there was no discount for young people, old people or families.

In this context, revenue management consists in deciding in real time to maintain the current fare class or to close it and move to the next one, resulting in a price increase. Coming back to a previous fare is impossible; thus, there are no last minute drops in ticket prices for trains that have still several empty seats. Also, revenue managers could decide to never open the first fare classes and begin to sell directly tickets in a higher fare class. Symmetrically, the last fare class may never be reached. In practice, revenue management was operated through a Computerized Reservation System (CRS). Before the beginning of sales, it fixes a seat allocation planning for all fare classes, using the history of purchases on past trains. During sales, the CRS uses the number of tickets sold up to now to make recommendations on the size of subsequent fare classes. Revenue management managers can nevertheless always intervene, both on the initial and on subsequent seat allocations, according to their experience on past trains.[7]

Finally, and crucially for our identification strategy, the revenue management did not use separate fare classes for a given train with several destinations. For instance, in a Paris-Toulouse train, the closure of the first fare class occurred exactly at the same moment for both Bordeaux and Toulouse. Hence, price changes of Paris-Bordeaux and Paris-Toulouse tickets happened exactly at the same time, for all trains. According to discussions with people in the revenue management department, this was to limit the number of decisions to be taken at each moment.

2.2 Data and descriptive statistics

We have data on iDTGV trains between May 2007 and March 2009 in economy class and for journeys from Paris to the rest of France. We first observe basic characteristics of the trains: all the stops, departure and arrival time, day of departure (e.g. May 2, 2008) and whether it corresponds to a peak time or not. We also observe the price grid used for that train for each fare class. For each route and type of period (peak time or off peak), there are a limited number of such grids, as they change these grids only a few times during the period we observe (e.g. 3 times for the Paris-Toulouse). We also observe the sales of each fare classes of all trains. On the other hand, we do not observe the purchasing dates, nor the opening moments of each fare class. For a given route, capacity is defined as the maximal number n such that for at least three trains, n seats were sold.[8]

Table 2 presents some descriptive statistics on our data. We observe a substantial

[7]Manager intervention in automatized revenue management also exists in other industries, e.g. hospitality industry (Cho et al., 2018).

[8]We use this definition (rather than the maximal number of seats sold across all trains of a given route) to take into account rare cases of overbooked trains. With this definition, we observe 5 cases of overbooking, over the $2,909$ trains of our dataset. Note that capacity can be assumed to be fixed for a given route because the number of coaches in economy class is fixed.

amount of price dispersion within trains. For instance on the Côte d'Azur line, the minimal price paid by consumers on average over the different trains (19.3€) was three times and a half lower than the average maximal price (68.4€). We also observe substantial variations on the average load across routes. While trains in Paris-Marseille were always nearly full, with an average load above 95%, this was far from being the case on the Côte basque line, with an average load of only 65.4%. This suggests that the actual pricing may not be fully optimal, at least for some routes.

Table 2: Descriptive statistics, economy class, from Paris

Route	Capacity	Avg Load	% final dest.	Prices Avg	Avg min.	Avg max.	max/min
Côte d'Azur	324	85.4%	81.5%	50.3	19.3	68.4	3.54
Marseille	324	95.5%	60.0%	49.5	19.0	70.5	3.71
Perpignan	324	88.6%	27.4%	50.0	20.2	72.6	3.59
Côte basque	350	65.4%	64.1%	37.3	19.7	53.3	2.71
Toulouse	350	87.3%	55.3%	43.6	19.4	67.2	3.46
Mulhouse	238	79.4%	24.1%	35.0	19.4	50.0	2.58

Notes: Avg min. and max. are the average of the minimal and maximal prices charged for each train, for the final destination. max/min is the ratio between the two previous columns.

3 Theoretical model and parameters of interest

3.1 Demand side

We consider a demand model close to that of McAfee and te Velde (2008). A train T is defined by its route $r(T)$ (e.g. Paris-Toulouse) and its day of departure (e.g. May 2, 2008). For each route r, we denote by a_r the intermediate destination and by b_r the final destination. To simplify notation and in the absence of ambiguity, we just denote the destinations of a train T by a and b instead of $a_{r(T)}$ and $b_{r(T)}$. For any train T, tickets are sold between the normalized dates $t = 0$ and $t = 1$. We denote the fare classes by $k \in \{1, ..., K\}$. Within fare class k, tickets for train T and destination $d \in \{a_{r(T)}, b_{r(T)}\}$ are sold at price p_{dkT}. We recall that p_{dkT} belongs to a grid of K

prices that remains fixed for several months and depends only on the destination d and whether the train leaves at a peak time or not. Finally, we denote by $V_{dT}(A, B)$ the number of consumers arriving during the subset A of the time interval $[0, 1]$, and with a valuation belonging to the subset B of $[0, \infty)$. Similarly, let $D_{dT}(t, t'; p_d)$ denote the demand for destination d in train T between dates t and t' (with $(t, t') \in [0, 1]^2$) when the price is constant and equal to p_d. Then $D_{dT}(t, t'; p_d) = V_{dT}([t, t'), [p_d, \infty))$. We then assume the following condition.

Assumption 1 *(Consumers' demand) For all T and $d \in \{a, b\}$, there exists $\varepsilon > 1$ and a random process $b_T(.)$ on $[0, 1]$, continuous and satisfying* $\min_{u \in [0,1]} b_T(u) > 0$ *almost surely, such that conditional on ξ_{dT} and $b_T(.)$:*

1. *V_{dT} is a Poisson process with intensity $I_{dT}(t, p) = \xi_{dT} b_T(t) \varepsilon p^{-1-\varepsilon}$ for $(t, p) \in [0, 1] \times [0, \infty)$. Without loss of generality, we let $\int_0^1 b_T(u)du = 1$.*
2. *V_{aT} and V_{bT} are independent.*

The term ξ_{dT} captures train-destination specific overall demand shocks. For instance, demand to Cannes may increase a lot during the Cannes Film Festival. The term $b_T(t)$ describes the pattern of consumers' arrival time for train T. We do not make any restriction hereafter on this function, nor do we impose it to be constant from one train to another. On the other hand, we impose that the intensity of V_{dT} takes a multiplicative form. This form has three implications. First, we assume that the arrival of consumers for destinations a and b have the same shape, as they are just shifted by a multiplicative destination-train specific constant ξ_{dT}. This condition can be tested, an important point on which we come back in Section 4.2 below. Second, we impose a specific functional dependence in p, of the form $p^{-1-\varepsilon}$. This particular form is not essential for our identification strategy. We do have to impose a parametric form, on the other hand, given that prices only take a few different values. When restricted to $[p_0, \infty]$ for any $p_0 > 0$, the intensity we consider corresponds to consumers' valuation following a Pareto distribution with a parameter equal to ε.

Finally, by considering a multiplicative form ($I_{dT}(t, p) \propto b_T(t) \times p^{-1-\varepsilon}$), we assume that the valuation of consumers does not evolve over time. In particular, Assumption

1 implies that the demand for destination d on the time interval $[t_1, t_2]$ satisfies

$$D_{dT}(t_1, t_2; p)|b_T(.) \sim \mathcal{P}\left(\xi_{dT} p^{-\varepsilon} \int_{t_1}^{t_2} b_T(u)du\right).$$

Thus, as McAfee and te Velde (2008), we assume that the price elasticity does not evolve over time. This assumption could be relaxed with more detailed data. We believe it is reasonable in our context where purchasers of the economy class tickets of these trains are already quite homogenous. Nonetheless, we test it and consider an extended model allowing for time-varying elasticities in Section 5.3 below.

Assumption 1 together with a supply-side restriction (Assumption 3) turns out to be sufficient to identify ε, see Point 1 of Theorem 4.2 below. To further identify the distribution of (ξ_{aT}, ξ_{bT}), we consider the next assumption. Hereafter, W_T denotes a vector of observed characteristics of train T (e.g., whether the train operates on a rush hour or not) and X_{dT} denotes a vector of observed characteristics of destination d served by train T, e.g., the travel time from Paris to d by train T.

Assumption 2 *For $d = \{a, b\}$, ξ_{dT} satisfies:*

(i). $\xi_{dT} = \exp\{X'_{dT}\beta_0\} g_0(W_T)\eta_{dT}$ *where η_{aT}, η_{bT} and (X_{aT}, X_{bT}, W_T) are independent.*

(ii). $\eta_{dT} \sim \Gamma(\lambda_{d0}, 1)$.

Assumption 2(i) specifies ξ_{dT} as the product of a function of X_{dT}, $g_0(W_T)$, and a remainder term η_{dT}. It restricts ξ_{aT} and ξ_{bT} to be dependent through the observed variables (X_{aT}, X_{bT}, W_T), rather than (η_{aT}, η_{bT}). This is plausible as long as one includes sufficient controls in (X_{aT}, X_{bT}, W_T). Importantly, we leave the function $g_0(.)$, which determines how train-specific characteristics affect the demand, unrestricted. Assumption 2(ii) imposes that conditional on X_{dT}, $g_0(W_T)$, ξ_{dT} follows a gamma distribution. Since we include a d-specific constant term in X_{dT}, we can normalize the scale parameter of the gamma distribution to 1. As detailed below, the assumption of a gamma distribution does not matter for identification. It is rather made for computational reasons: that the gamma and Poisson distributions are conjugate makes it possible to simplify the computation of counterfactual revenues under incomplete information, see Online Appendix D for more details. We also consider log-normality as a robustness check below, but without computing all counterfactual revenues, then.

Note that in (i), we made the simplifying assumption that trains only had two destinations, an intermediate a and a final one b. But recall from Table 1 that most of them serve more than just two cities, so a or b actually correspond to more than one city. If so, we modify (i) by assuming that

$$\xi_{dT} = \left[\sum_{c \in d} \exp(X'_{cT}\beta_0)\right] g_0(W_T)\eta_{dT}, \tag{1}$$

where c is an index for cities belonging to either a or b. For instance, in a train to Côte d'Azur, c corresponds to Avignon for destination a whereas c includes Cannes, Saint-Raphaël and Nice, see again Table 1.[9]

3.2 Supply side

We now formalize the features of revenue management already discussed in Section 2.1. First, recall that the revenue management is operated at a route level (e.g. Paris-Toulouse) rather than for each destination of this route (e.g. Paris-Bordeaux and Paris-Toulouse for the route Paris-Toulouse). We thus make the following assumption.

Assumption 3 *(revenue management at the route level) The opening time of fare class $k \in \{1, ..., K\}$, τ_k, is a stopping time with respect to the process $t \mapsto N_{aT}(t) + N_{bT}(t)$, where $N_{dT}(t)$ is the number of purchases for d made before t.*

Assumption 3 states that the decision of opening a new fare class depends only on past total purchases, rather than on the repartition between purchases for a and for b. Such an assumption is fully in line with the fact that a single fare class is used for the two destinations of each route. It was also confirmed by discussions we had with the revenue management department.

Our second assumption on the supply side is a weak optimality condition for the firm. To introduce it, let $R_T(p_a, p_b)$ denote the maximal revenue for train T under

[9]Note, on the other hand, that all cities $c \in d$ are priced equally, so we do not need to take into account price variations between cities.

a uniform pricing of (p_a, p_b) for destinations a and b respectively. This maximal revenue is obtained by considering the optimal quotas C_{aT} and C_{bT} of tickets sold for destinations a and b respectively, with $C_{aT} + C_{bT} = C_T$, the total capacity of train T (the exact formula of $R_T(p_a, p_b)$ is displayed in Equation (6) below). Let also p_{dkT} denote the price in train T and fare class $k \in \{1, ..., K\}$ for destination $d \in \{a, b\}$. The weak optimality condition we consider is the following:

Assumption 4 *(Weak optimality of actual revenue management) We have*

$$\max_{k=1,\dots,K} \mathbb{E}\left[R_T(p_{akT}, p_{bkT})|W_T\right] \leq \mathbb{E}\left[R_T^{obs}|W_T\right]. \tag{2}$$

By conditioning on W_T, which only includes coarse proxies of the true demand, we allow for the possibility that revenue managers use limited information for their pricing strategy. In reality, it seems credible that they have access to additional signals on the true demand for a specific train. For instance, they could use the past number of purchases in each fare class on previous years for the same exact train. If so, we would expect that Inequalities (2) would also hold conditional on this information.

Importantly, Assumption 4 does not imply that the revenue management performs better than the optimal uniform pricing, because we only impose that observed revenues exceed any uniform pricing strategy that is constrained to the grid of the 12 predetermined prices. In other words, we simply assume that observed revenues are on average higher than those one would have obtained by sticking from $t = 0$ to $t = 1$ to one of the fare class.

Moreover, we do not impose any optimality with respect to all dynamic strategies. We refrain to do so for several reasons. First, such an assumption would conflict with our very objective to quantify the gains or losses of the actual revenue management, compared to alternative scenarios. By definition, assuming a strong form of optimality would result in gains against most simpler pricing strategies. Second and related to the first point, it seems very restrictive in our setting to assume that the optimal dynamic strategy was adopted. As discussed in Section 2.1, the revenue management applied simplified rules (increasing fares from 12 predetermined fare classes), which can at best approach the optimal solution. Moreover, seat allocation decisions were also subject to the manager's manual intervention, which could be a source of suboptimality.[10] Further, computing the optimal dynamic strategy under the simplified rules is still a very complicated dynamic programming problem. While Feng and Xiao (2000) have proposed an algorithm for computing the solution for a homogeneous Poisson process, little has been done so far for the non-homogeneous case, to our knowledge. Finally, given that iDTGV has been merely created in 2004, we can doubt that it perfectly knows the demand parameters, and in particular all destination-train-specific effects ξ_{dT}.

3.3 Parameters of interest

We aim at comparing the current revenues with several counterfactual revenues, depending on the type of revenue management and the information the firm has access to. We consider several possible pricing strategies, from the most basic to the most sophisticated ones. The first, uniform pricing, simply corresponds to fixing the price of each route in a given train once and for all. We let R_u denote optimal counterfactual revenues, averaged over all trains, under this pricing regime. At the other extreme, in "full" dynamic pricing, prices can be changed any time. R_f then corresponds to optimal counterfactual revenues in this set-up. We also study pricing strategies, called stopping-time strategies hereafter, where prices can be changed only after a ticket is sold. The corresponding optimal revenues are then R_s. Finally, we consider constrained stopping-time strategies close to what was implemented in practice, by assuming that only M number of fares, or M increasing fares, are allowed. The corresponding optimal revenues are denoted by R_{sM} and R_{sM+}, respectively. To compute these counterfactual revenues, we maintain Assumption 1. This means that for pricing strategies where prices are allowed to decrease, we rule out any forecast of such price decreases by consumers.

Hereafter, we consider two scenarios in terms of information available to the revenue managers.

1. (Complete information) Revenue managers fully know the expected demand for each train. Thus, they observe ε, $b_T(.)$, ξ_{aT} and ξ_{bT} for each train T;

[10] See Cho et al. (2018, 2019); Phillips (2021) for evidence of suboptimality due to human management.

2. (Incomplete information) Revenue managers observe ε, $b_T(.)$, (X_{aT}, X_{bT}, W_T) but only $f_{\xi_{aT},\xi_{bT}|X_{aT},X_{bT},W_T}$. As time goes by, revenue managers update their information on (ξ_{aT}, ξ_{bT}) according to Bayes' rule.

The complete information case should be seen as a benchmark. It is useful in particular to quantify the value of information and contrast the gains of revenue management in complete and incomplete information set-ups. The case of incomplete information is probably more realistic. In this scenario, revenue managers know, for each train, the pattern of consumers' arrival over time ($b_T(.)$) but does not know exactly the aggregate demand for each destination (ξ_{aT} and ξ_{bT}). The assumption that $b_T(.)$ is known makes especially sense if $b_T(.)$ does not depend on T, in which case revenue managers can have learned from previous trains how consumers arrive through time.

If the scenario of incomplete information holds in practice, the differences between the counterfactual revenues and the observed ones can be interpreted as the potential gains or losses of the optimal revenue management under different constraints compared to the actual ones. Hereafter, we use the exponents c and i to specify the two information set-ups. Hence, R_u^c denotes for instance the counterfactual optimal revenue under uniform pricing and complete information.

In all the counterfactual scenarios, we consider a separate pricing strategy for destinations a and b, contrary to the actual practice. On the other hand, for computational reasons, we cannot consider the fully optimal pricing strategies for the two destinations. Without further restrictions, the state space is large: the optimal strategy at any time t depends on the remaining seats for both destination. To reduce this state space, we fix ex ante the total number of seats available for stops a (C_{aT}, say) and thus to b ($C_{bT} = C_T - C_{aT}$, with C_T the total number of seats in train T). Then, depending on the scenario we consider, we either consider the optimal pre-allocation C_{aT}, or fix it so that C_{aT} matches the observed average sales for a. In any case, fixing C_{aT} allows us to solve the optimization problem separably for each destination (given the independence of η_{aT} and η_{bT} imposed in Assumption 2(i)), rather than jointly. This greatly reduces the computational burden of the optimization problem. For this reason, our results below may be seen as lower bounds on the fully optimal counterfactual revenues. This actually reinforces some of our conclusions below. Also, we compare in Section 5.3 below the revenues under uniform pricing with and without pre-allocation, and do not find important differences between the two.

4 Identification and estimation

In this section, we first clarify which parameters of the demand function are needed to recover the average revenues under the counterfactual scenarios described above. We also describe challenges for the identification of these parameters. Next, we show how price elasticity and relative demand effects can be identified. We then describe the partial identification of the distribution of train effects and counterfactual revenues. Finally, we show how to perform inference on the parameters of interest. The asymptotic framework we consider below is obtained by letting the number of trains tend to infinity (recall that in our application, we observe 2,909 trains).

4.1 A first result and challenges

The following theorem clarifies which parameters of the demand are required to identify all the counterfactual revenues we consider.

Theorem 4.1 *Suppose that Assumptions 1-2 hold. Then, R_r^I is a function of the distribution of $(\xi_{aT}, \xi_{bT}, X_{aT}, X_{bT}, W_T)$ and ε, for $I \in \{c, i\}$ and $r \in \{u, f, s, sM, sM+\}$.*

We obtain the result by constructing new Poisson demand processes $\widetilde{V}_{aT}$ and $\widetilde{V}_{bT}$ with the same parameters as the true ones V_{aT} and V_{bT}, except that they are homogeneous: $b_T(.)$ is replaced by $\widetilde{b}_T(.) = 1$. We prove that the optimal revenues are the same for these new demand processes as for the original ones. This shows that the optimal revenues depend on $I_{dT}(.,.)$ only through ξ_{aT} and ξ_{bT}: it does not matter whether consumers arrive early or late, as long as on average, the same number of consumers eventually arrive. The result holds because basically, all the constraints on pricing we consider are independent of time. In this sense, Theorem 4.1 holds beyond the specific scenarios we consider here. But it would fail if time constraints were imposed on the pricing strategies, for instance if a limit on the number of price changes occuring before a given date $t^* < 1$ was set.

Theorem 4.1 is crucial in our context with no information on purchasing dates. In the absence of such information, there is no way to recover $b_T(.)$. Instead, we only have to recover the price elasticity ε and the conditional distribution of destination-train-specific effects (ξ_{aT}, ξ_{bT}) to identify the counterfactual revenues.

We do not specify here the exact forms of the counterfactual revenues, as they do not have closed forms. However, we can obtain them by induction, using the Bellman equations associated with the optimal strategies and solving some differential equations. In the incomplete information case, the gamma specification in Assumption 2(ii) is helpful for that purpose, as the gamma distribution is a conjugate prior for Poisson likelihood. The induction formulas are given in Online Appendix D.

McAfee and te Velde (2008) obtains a similar result as Theorem 4.1 for the "full" dynamic pricing strategy under complete information and a similar demand model. We extend their results in two directions. First, we consider other types of pricing strategies, and in particular possibly constrained stopping-time strategies, which are very common in practice and correspond to the actual revenue management. Second, we also show a similar result in an incomplete information set-up.

Now, we face two main issues for recovering the demand parameters. First, demand is actually unobserved; only bounds on it can be obtained. Let n_{dkT} denote the number of sales for train T, fare class $k \in \{1, ..., K\}$ and destination $d \in \{a, b\}$. Then

$$D_{dT}(p_{dkT}) \geq D_{dT}(\tau_{k,T}, \tau_{k+1,T}; p_{dkT}) = n_{dkT},$$

where $\tau_{k,T}$ is the (random) time at which the kth fare class opens, which we do not observe. Hence, without further assumptions, we only observe a crude lower bound on the total demand at price p_{dkT}. This point was already made in similar contexts by Swan (1990), Lee (1990), and Stefanescu (2012).

The second issue we face is the absence of usual instruments for prices. Prices only vary within the grid specified by revenue managers, and to our knowledge, fare classes did not close for exogenous reasons unrelated to demand. In other words, there is no exogenous variations of prices in our context. The bottom line is that usual strategies to identify the demand function do not apply here.

We now show that despite these limitations, it is possible, under Assumptions 1-3, to point or partially identify the parameters $(\theta_0, g_0(.))$, where $\theta_0 := (\varepsilon, \beta_0, \lambda_{a0}, \lambda_{b0})$ and $\beta_0, \lambda_{a0}, \lambda_{b0}$ and $g_0(.)$ are defined in Assumption 2. Then, in view of Theorem 4.1, we obtain bounds on the counterfactual revenues. We proceed in two steps hereafter, by first showing point identification of θ_0 and then partial identification of $g_0(.)$.

4.2 Point identification of θ_0

We first identify ε by exploiting variations in the relative prices p_{bkT}/p_{akT} between the two destinations and from one fare class to another. We start from $n_{dkT} = D_{dT}(\tau_k, \tau_{k+1}; p_{dkT})$. For the sake of exposition, let us first assume that τ_k and τ_{k+1} are deterministic. Then, by Assumption 1, $D_{aT}(\tau_k, \tau_{k+1}; p_{ak})$ and $D_{bT}(\tau_k, \tau_{k+1}; p_{bk})$ are independent conditional on ξ_{aT}, ξ_{bT} and $\int_{\tau_k}^{\tau_{k+1}} b_T(u)du$. Moreover, they both follow Poisson distributions. As a result,

$$n_{bkT}|n_{akT} + n_{bkT} = n, \xi_{aT}, \xi_{bT} \sim \text{Binomial}\left(n, \Lambda(\ln(\xi_{bT}/\xi_{aT}) - \varepsilon \ln(p_{bkT}/p_{akT}))\right), \quad (3)$$

where $\Lambda(x) = 1/(1 + \exp(-x))$. The term $\ln(\xi_{bT}/\xi_{aT})$ may be seen as a train fixed effect. Hence, this model boils down to a fixed effect logit model, and ε is identified as long as there are variations through fare classes k in the relative prices p_{bkT}/p_{akT}. In the data, we do observe such variations. In Paris-Toulouse for instance, p_{bkT}/p_{akT} vary from 1 for $k = 1$ to 1.18 for $k = 12$. Then, if we add Assumption 2(i), we can intuitively identify β_0 and λ_0 from the fact that we have a random effect logistic model. Note that $g_0(W_T)$ is canceled out in the ratio ξ_{bT}/ξ_{aT}. In Section 4.3 below, we use further arguments to partially identify this function.

To obtain (3), we assumed that the stopping times $(\tau_k)_{k=1,...,K}$ were fixed, which is unrealistic. Nonetheless, the following result shows that (3), and then identification of θ_0, still holds provided that these stopping times satisfy Assumption 3.

Theorem 4.2 *Suppose that Assumptions 1 and 3 hold and that with positive probability, $k \mapsto p_{bkT}/p_{akT}$ is not constant. Then,*

1. *Equation* (3) *holds and ε is point identified;*
2. *If Assumptions 2(i) and 5 in the appendix further hold, β_0 and the distribution of η_{bT}/η_{bT} are identified.*

Two remarks are in order. First, Equation (3) does not hold for any possible random stopping times. We can easily build counterexamples by making $(\tau_k)_{k=1,...,K}$ depend solely on $N_{aT}(.)$, for instance. Such situations are however ruled out by Assumption 3. Under this condition, intuitively, the stopping times will be independent of the proportion of consumers buying tickets for a (versus b). Second, we actually prove the nonparametric identification of η_{bT}/η_{bT}. This implies the identification of λ_0 under Assumption 2(ii). It also shows that imposing this latter condition is not necessary for identification. As mentioned above, it solely matters for the computation of counterfactual revenues.

Beyond the identification of θ_0, Equation (3) can be the basis of testing some of the conditions we have imposed. First, the separability between $b_T(.)$ and ξ_{dT} in Assumption 1 implies that if $p_{bkT} = p_{akT}$ for several fare classes k, we should observe similar proportions $n_{bkT}/(n_{akT} + n_{bkT})$ for the corresponding k. Second, we can also test for the fact that price elasticities do not evolve over time, by considering more general specifications than (3). Third, we have imposed so far that the price elasticity was constant for all routes. We made this restriction for parsimony and consistency, because several routes share common origin-destination sections (e.g. Paris-Toulouse and Paris-Côte basque share the Paris-Bordeaux section). But we can allow for variations according to the day and hour of departure and according to groups of routes sharing the same sections. We consider all these extensions and robustness checks in Sections 5.1 and 5.3 below.

4.3 Partial identification of $g_0(W_T)$

To partially identify $g_0(W_T)$, which corresponds to the train-specific effect in ξ_{dT}, we build moment inequalities based on consumers' rationality (Assumption 1.1) and weak optimality of the actual revenue management (Assumption 4).

Consumers' rationality First, by Assumption 1.1, all consumers who bought a ticket for d at price p_{djT} for $j \geq k$ would have also bought it at price p_{dkT}. Therefore, for all $k = 1, ..., K$ and $d \in \{a, b\}$,

$$D_{dT}(p_{dkT}; g_0(W_T), X_{dT}) \geq \sum_{j=k}^{K} n_{djT},$$

where we now index total demand $D_{dT}(p_{dk})$ by $g_0(W_T)$ and X_{dT}. Let C_T denote the capacity of train T. Then we also have $C_T \geq \sum_{j=k}^{K} n_{djT}$. Combining these inequalities and integrating conditional on W_T, we obtain, for all $k = 1, ..., K$ and $d \in \{a, b\}$,

$$\mathbb{E}\left[\sum_{j=k}^{K} n_{djT} - C_T \wedge D_{dT}(p_{dkT}; g_0(W_T), X_{dT})\middle| W_T\right] \leq 0. \tag{4}$$

We assume hereafter that X_{dT} is a deterministic function of W_T. This holds in our context where W_T includes indicator of routes and X_{dT} includes time-invariant destination variables and interactions between such variables and W_T. Then, the function $g \mapsto \mathbb{E}[C_T \wedge D_{dT}(p_{dk}; g, X_{dT})|W_T]$ is strictly increasing. Denoting by $Q_k^{-1}(.; W_T, \theta_0)$ its inverse, we get

$$g_0(W_T) \geq Q_k^{-1}\left(\mathbb{E}\left[\sum_{j=k}^{K} n_{djT}\middle| W_T\right]; W_T, \theta_0\right).$$

Then, we obtain a lower bound for $g_0(W_T)$:

$$g_0(W_T) \geq g_0^L(W_T) := \max_{\substack{d=a,b \\ k=1,...,12}} \left\{Q_k^{-1}\left(\mathbb{E}\left[\sum_{j=k}^{K} n_{djT}\middle| W_T\right]; W_T, \theta_0\right)\right\}. \tag{5}$$

While Q_k^{-1} does not have a closed form, we can compute it easily through simulations.

Weak optimality condition We now rely on Assumption 4 to form additional moment inequalities. To exploit them, note that under Assumptions 1-2, we have (see Appendix D.2.2, section "uniform pricing", for details)

$$\mathbb{E}[R_T(p_a, p_b)|W_T] = \max_{\substack{(C_{aT}, C_{bT}): \\ C_{aT}+C_{bT}=C_T}} \left\{\sum_{d \in \{a,b\}} p_d \int_0^{\infty} \mathbb{E}\left[D(\exp\{X_{dT}'\beta_0\} p_d^{-\varepsilon} g_0(W_T) z) \wedge C_{dT}|W_T\right] \times g_{\lambda_{d0},1}(z) dz\right\}, \tag{6}$$

where $D(u) \sim \mathcal{P}(u)$, $g_{\lambda_{d0},1}$ is the density of a $\Gamma(\lambda_{d0}, 1)$, and C_{dT} is the total number of seats allocated to destination d. As a result,

$$R(g_0(W_T); X_{aT}, X_{bT}, \theta_0) := \max_{k=1,...,K} \mathbb{E}\left[R_T(p_{akT}, p_{bkT})|W_T\right]$$

is an identified function. Note that again, our notation reflects that (X_{aT}, X_{bT}) is a deterministic function of W_T. Hence, the weak optimality condition (2) rewrites as

$$R(g_0(W_T); X_{aT}, X_{bT}, \theta_0) \leq \mathbb{E}\left[R_T^{\text{obs}}|W_T\right]. \tag{7}$$

The function $R(.; X_{aT}, X_{bT}, \theta_0)$ is strictly increasing. Denoting by $R^{-1}(.; X_{aT}, X_{bT}, \theta_0)$ its inverse, we obtain the following upper bound for $g_0(W_T)$:

$$g_0(W_T) \leq g_0^U(W_T) = R^{-1}\left(\mathbb{E}\left[R_T^{\text{obs}}|W_T\right]; X_{aT}, X_{bT}, \theta_0\right) \tag{8}$$

4.4 Partial identification of counterfactual revenues

As shown by Theorem 4.1, R_r^I ($I \in \{c, u\}$) is a function of the distribution of $(\xi_{aT}, \xi_{bT}, W_T)$ and price elasticity ε. Further, under Assumption 2, given W_T and a pre-allocation (C_{aT}, C_{bT}), R_r^I has the following form (see Appendix D):[11]

$$R_r^I(W_T, C_{aT}, C_{bT}) = \sum_{d=a,b} \alpha_r^I(C_{dT}, \varepsilon, \lambda_{d0}) \exp\{X'_{dT}\beta_0/\varepsilon\} g_0(W_T)^{1/\varepsilon},$$

for some non-random term $\alpha_r^I(C_{dT}, \varepsilon, \lambda_{d0})$. Then, using the bounds of $g_0(W_T)$ in (5) and (8), we obtain lower and upper bounds for $R_r^I(W_T, C_{aT}, C_{bT})$ as:

$$\left[g_0^L(W_T)^{1/\varepsilon}, g_0^U(W_T)^{1/\varepsilon}\right] \times \sum_{d=a,b} \alpha_r^I(C_{dT}, \varepsilon, \lambda_{d0}) \exp\{X'_{dT}\beta_0/\varepsilon\}. \tag{9}$$

Bounds on R_r^I then follow by averaging (9) over trains:

$$\mathbb{E}\left[\left[g_0^L(W_T)^{1/\varepsilon}, g_0^U(W_T)^{1/\varepsilon}\right] \times \sum_{d=a,b} \alpha_r^I(C_{dT}, \varepsilon, \lambda_{d0}) \exp\{X'_{dT}\beta_0/\varepsilon\}\right]. \tag{10}$$

We also consider below ratios of counterfactual revenues. Given what precedes, such ratios r_0 satisfy

$$r_0 = \frac{\mathbb{E}[f_1(U_T)g_0(W_T)^{1/\varepsilon}]}{\mathbb{E}[f_2(U_T)g_0(W_T)^{1/\varepsilon}]},$$

for two identified, positive functions f_1 and f_2. Let $\mathcal{R}$ denote the identified set on r_0. Then one can show that $\mathcal{R}$ is an interval $[\underline{r}, \overline{r}]$, where $\overline{r}$ and $\underline{r}$ are defined as the unique solutions of

$$\mathbb{E}\left[g_0^L(W_T)^{1/\varepsilon}(f_1(U_T) - \overline{r} f_2(U_T)) + (g_0^U(W_T)^{1/\varepsilon} - g_0^L(W_T)^{1/\varepsilon})(f_1(U_T) - \overline{r} f_2(U_T))_+\right] = 0,$$
$$\mathbb{E}\left[g_0^U(W_T)^{1/\varepsilon}(f_1(U_T) - \underline{r} f_2(U_T)) + (g_0^L(W_T)^{1/\varepsilon} - g_0^U(W_T)^{1/\varepsilon})(f_1(U_T) - \underline{r} f_2(U_T))_+\right] = 0.$$

4.5 Estimation and inference

We estimate θ_0 as follows. Let $Y_{jkT} = 1$ if seat j in fare class k for train T is sold to a, $Y_{jkT} = 0$ otherwise. By (3), we have

$$\Pr(Y_{jkT} = 1 | \xi_{aT}, \xi_{bT}) = \Lambda\left(\ln(\xi_{bT}/\xi_{aT}) - \varepsilon \ln(p_{bkT}/p_{akT})\right),$$

[11] The only exception is for revenues under uniform pricing with prices constrained to belong to the grid, for which the form is more complicated. Specifically, it corresponds to the maximum over the grid of the revenue displayed in (6). Nonetheless, we can still simply obtain bounds on these revenues using the monotonicity of the right-hand side of (6) with respect to $g_0(W_T)$.

and the $(Y_{jkT})_{j=1,...,n_{kT}}$ (with $n_{kT} := n_{akT} + n_{bkT}$) are independent. Thus, we can estimate ε and $\ln(\xi_{bT}/\xi_{aT})$ by maximizing the likelihood of a logit model including train fixed effects. Because the number of sales for each train is large (usually above 250), the bias related to the estimation of these fixed effects is expected to be negligible. Second, under Assumption 2 (with the equality in (i) replaced by (1) to account for multiple cities in each $d \in \{a, b\}$),

$$\ln(\xi_{bT}/\xi_{aT}) = \ln\left[\frac{\sum_{c\in b}\exp(X'_{cT}\beta_0)}{\sum_{c\in a}\exp(X'_{cT}\beta_0)}\right] + \ln\left(\frac{\eta_{bT}}{\eta_{aT}}\right), \quad \eta_{bT}/\eta_{aT} \perp\!\!\!\perp (X_{cT})_c.$$

Then, we estimate β_0 by nonlinear least squares, replacing $\ln(\xi_{bT}/\xi_{aT})$ by its estimator. Finally, we estimate λ_0 by maximum likelihood on the sample $(\ln(\widehat{\eta_{bT}/\eta_{aT}}))_T$, with

$$\ln(\widehat{\eta_{bT}/\eta_{aT}}) = \ln(\widehat{\xi_{bT}/\xi_{aT}}) - \ln\left[\frac{\sum_{c\in b}\exp(X'_{cT}\widehat{\beta})}{\sum_{c\in a}\exp(X'_{cT}\widehat{\beta})}\right].$$

In principle, we could directly estimate θ_0 by maximum likelihood, as under Assumptions 1-2, the distribution of $(Y_{jkT})_{j=1,...,n_{kT},k=1,...,K}$ is fully parametric. We do not adopt this method for two reasons. First, the estimators of ε and λ_0 would be sensitive to the parametric specification on (η_{aT}, η_{bT}). Second, the corresponding estimator is much more complicated to compute, something turning out to be important when considering inference based on the bootstrap.

Next, we estimate the lower and upper bounds on $g_0(W_T)$ by the empirical counterparts of (5) and (8), where the conditional expectations $\mathbb{E}(.|W_T)$ are replaced by empirical means (as W_T is discrete in our specification, see below for its full description). Finally, we estimate bounds on R_r^I by the empirical counterpart of (10).

As estimation involves multiple steps, we rely on the bootstrap for inference. We compute confidence intervals on counterfactual revenues with nominal levels of $1-\alpha$ as follows. The lower bound corresponds to the $\alpha/2$-th quantile of the bootstrapped lower bound in (10), while the upper bound corresponds to the $1-\alpha/2$-th quantile of the bootstrapped upper bound in (10). This ensures an asymptotic coverage of at least $1-\alpha$, whether the parameter is point or partially identified.

5 Results

5.1 Demand estimation

We first consider the estimation of the price elasticity ($-\varepsilon$), the coefficients of destination-train specific effects (β_0), and the parameters of Gamma distribution λ_0. The variables we include in W_T are route dummies, time dummies for the year and month of the train, whether it occurs during the weekend, on public holidays, on school holidays and whether the departure time is during rush hour. Regarding the variables X_{dT} or, to be more precise, X_{cT} where c denotes a city (see our discussion around Equation (1)), we include travel time to c by train T, its square, city-specific effects X_c (namely, the population of the urban area of c and whether c is a regional capital) and all interactions $X_{cj} \times W_{Tk}$ for all components X_{cj} and W_{Tk} of the vectors X_c and W_T, respectively.

The estimates of price elasticities are displayed in the top panel of Table 3. In Column I (our baseline specification), we assume a constant price elasticity across routes and trains and obtain a price elasticity of -4.04. This result is larger (in absolute value) than those in the literature on the transportation industry. We refer for instance to the meta-analysis by Jevons et al. (2005) and the studies of Wardman (1997), Wardman (2006) and Wardman et al. (2007), which point to price elasticities in the range $[-1.3; -2.2]$. Unlike ours, most of the studies rely on aggregated data. This is likely to bias upwards price-elasticity estimates, a point that we illustrate in Appendix B by running regressions based on our data aggregated at different levels.

The middle panel of Table 3 reports the estimates of the components of β_0 corresponding to the travel time and city-specific effects. The effect of the population size and travel time by train are as expected. Larger cities lead to higher demand and a longer travel time by train leads to a lower demand for train tickets. The effect of travel time may nonetheless be attenuated for long journeys, though the coefficient of the square of travel time is not significant.

The bottom panel of Table 3 reports the estimates of the parameters $(\lambda_{a0}, \lambda_{b0})$ of the gamma distribution. Intermediate destinations are estimated to have larger uncertainty on demand ($V(\eta_{dT}) = \lambda_{d0}$ under the gamma specification), though the difference between the two is not statistically significant.

Table 3: Estimates of $(\varepsilon, \beta_0, \lambda_0)$

	Binomial model		Multinomial model	
	I	II	III	IV
Price elasticity (ε)				
Constant	4.04 (0.22)	6.96 (0.57)	4.04 (0.22)	6.96 (0.38)
Southwest		−2.09 (0.54)		−2.09 (0.18)
Weekend/national holidays		−2.46 (0.51)		−2.46 (0.17)
Peak hour		0.37 (0.48)		0.37 (0.15)
Destination effects				
Population (in M. inhabitants)	2.23 (0.20)	2.24 (0.20)	2.17 (0.19)	2.18 (0.19)
Regional capital	0.20 (0.19)	0.20 (0.19)	0.17 (0.19)	0.17 (0.20)
Travel time by train (in hours)	−2.07 (0.11)	−2.11 (0.11)	−1.60 (0.11)	−1.64 (0.11)
Travel time by train, squared	0.34 (0.66)	0.35 (0.66)	0.28 (0.64)	0.29 (0.65)
Gamma distributions				
λ_{a0} (intermediate)	3.63 (1.01)	3.63 (1.00)	3.93 (1.07)	3.93 (1.06)
λ_{b0} (final)	2.62 (0.40)	2.62 (0.40)	2.79 (0.41)	2.78 (0.41)
Control for $X_d \times W_T$	Yes	Yes	Yes	Yes
R^2 of the reg. of $\ln(\xi_{bT}/\xi_{aT})$	0.502	0.506	0.521	0.525

Notes: The total number of trains is 2,909. In Columns I and II, with all fare classes, the total number of observations (fare classes × trains) is 21,988. In Columns III and IV, the total number of observations (fare classes × trains) is 34,908. Southwest correspond to the lines to Côte Basque, Toulouse and Perpignan. Standard errors (under parentheses) are calculated using the bootstrap with 500 re-sampled datasets.

In Column II, we estimate the demand model by allowing price elasticity to vary across routes and trains. We find that travelers of routes from Paris to the southwest of France (namely, the routes to Côte basque, Toulouse and Perpignan) are less price-sensitive to price than those of other routes. Travelers on weekend or national holidays have a smaller price elasticity (in absolute value) than those on other days. On the other hand, once controlling for weekend and national holidays, individuals traveling during peak hours appear to have a similar elasticity to the others.

For several routes, there are actually multiple intermediate or final destinations. If Assumptions 1 and 3 hold, Theorem 4.2 implies that the joint distribution of the purchases for these multiple destinations, conditional on the total number of purchases on the train, is multinomial, rather than binomial in case the purchases for the intermediate or final stops are aggregated. We re-estimate the demand models corresponding to Columns I and II using a multinomial model. The results are displayed in Columns III and IV, respectively. The resulting price elasticities are almost identical to those obtained before. The destination effects and estimates of λ_0 are also very similar.

5.2 Counterfactual revenues

We now turn to the counterfactual revenues under different pricing strategies, namely uniform, stopping-time, and full dynamic pricing. For counterfactual revenues R_r^I with $r \in \{u, f, s\}$ and $I \in \{c, i\}$, we simulate the revenue with the optimally pre-allocated numbers of available seats for intermediate and final stops; for R_r^I with $r \in \{sM, sM+\}$ and $I \in \{c, i\}$, i.e. stopping-time pricing strategy with M (increasing) fares, we fix the pre-allocated number of available seats for intermediate stop a, C_{aT}, to be equal to the average number of seats sold for a among all the trains operated on the given route. We do this, rather than finding the optimal value of C_{aT}, for computational reasons. Moreover, for the other pricing strategies ($r \notin \{sM, sM+\}$), the revenues obtained this way secures at least 99% of the revenue based on the optimal pre-allocation, so we expect very little effect of considering this specific pre-allocation.

Table 4 summarizes the set estimates of counterfactual revenues averaging over all routes based on Column I in Table 3 – we discuss the results based on Column II in Section 5.3 below. When possible, we indicated the 95% confidence intervals on the set.[12] Below, we organize our discussion of the results along different themes.

[12] Computing the set estimates of counterfactual revenues can be costly, as it involves the terms α_r^I, which are only defined by induction and thus can take time to be obtained. For instance, computing the set estimates corresponding to Line s.3 takes us 77 hours. For 500 bootstrap replications executed on 10 cores, this would mean 160 days of computational time.

Table 4: Revenues under counterfactual pricing strategies, average over lines

Scenario	Point or set estimate (in K€)
Observed pricing strategy	12.21 [12.05,12.36]
Uniform pricing strategy	
u.1 Incomplete information, constrained prices	$[11.15, 12.21]$ [10.78,12.41]
u.2 Incomplete information, unconstrained prices	$[11.29, 12.31]$ [11.04,12.54]
u.3 Complete information, constrained prices	$[12.56, 13.86]$ [12.31,14.25]
u.4 Complete information, unconstrained prices	$[13.23, 14.42]$ [13.04,14.87]
Stopping-time pricing strategy	
s.1 Incomplete information, 2 increasing fares	$[12.66, 13.90]$
s.2 Incomplete information, 2 fares	$[12.85, 14.11]$
s.3 Incomplete information, 12 increasing fares	$[13.10, 14.39]$
s.4 Incomplete information, 12 fares	$[13.27, 14.57]$
s.5 Incomplete information	$[13.44, 14.66]$ [13.25,15.10]
s.6 Complete information, 2 increasing fares	$[13.34, 14.54]$
s.7 Complete information, 2 fares	$[13.41, 14.62]$
s.8 Complete information, 12 increasing fares	$[13.39, 14.60]$
s.9 Complete information, 12 fares	$[13.47, 14.69]$
s.10 Complete information	$[13.48, 14.70]$ [13.29,15.15]
"Full" dynamic pricing strategy	
f.1 Incomplete information	$[13.47, 14.68]$ [13.28,15.13]
f.2 Complete information	$[13.50, 14.72]$ [13.31,15.17]

Notes: Estimated bounds on counterfactual revenues, with 95% confidence intervals below. These intervals are obtained with 500 bootstrap replications. With "constrained prices" (resp. "unconstrained prices"), optimization is conducted over the actual price grid (resp. over all positive real numbers).

How does the actual strategy compare to counterfactual pricing strategies? Recall that by Assumption 4, the actual strategy is supposed to be better than any uniform pricing strategy under incomplete information and with prices constrained to belong to the price grid. The gains are however moderate: they range between 0% and 9.5%. Then, we already cannot exclude that the actual strategy performs actually worse than the same uniform pricing strategy but with unconstrained prices (see Scenario u.2). In any case, the gains would be at most 8.1%. When turning to the most constrained dynamic pricing strategy, namely two fare classes and increasing prices, we observe a loss in revenue ranging between 3.6% and 12.2%. When we consider the same constraints as in the actual pricing strategy, namely 12 fare classes and increasing prices, we estimate a loss in between 6.8% and 15.1%.

Because of fixed pre-allocations for destinations a and b, the revenues in Table 4 are just lower bounds on the true, optimal revenues, which still reinforces our conclusions above. To get a sense on the quantitative effect of these pre-allocations, we simulate counterfactual revenues under unconstrained uniform pricing without pre-allocating capacities among intermediate and final destinations. The corresponding formulas are in Appendices D.1.2 and D.2.2, see the sections "uniform pricing" therein. In the complete information set-up, we obtain a set estimate of $[13.45, 14.68]$, corresponding to an increase in between 1.4% and 1.8% compared to Scenario u.4. In the incomplete information set-up, we obtain a higher gain of around 6%, with a set estimate of $[11.96, 14.68]$. This 6% might be the upper bound on possible gains from not imposing any pre-allocation, as one could expect that the effects of pre-allocation can be more easily mitigated with more flexible pricing strategies.

How can we explain the suboptimality of the actual strategy, in particular compared to the optimal strategies under similar pricing constraints? First, the initial seat allocation planning determined by the CRS may sometimes be far away from the optimal allocation under complete information. Then, revenue managers may fail to adjust enough this initial allocation. Second, in our counterfactuals, we have considered that revenue managers knew the true ε, the true effects of covariates, or the true $b_T(.)$. This may not be the case in reality. In any case, our results emphasize the importance of not imposing strong optimality conditions on the supply side.

Does it matter to have a fixed price grid? We look at this question by comparing the revenues obtained under optimal uniform strategies with prices either chosen optimally on $[0, \infty)$ or only within the actual price grid of the train under consideration. The effect of the grid is higher in the complete information set-up, with a gain of an unconstrained optimization roughly ranging in between 4% and 5.3%.[13] This is basically because demand is very high or very low for a few trains, in which case one would like to set a price above the maximal price, or below the minimal price of the grid. On the other hand, fixing the price grid has very small effects on revenues under incomplete information, with gains in between 0.8% and 1.3%.

Does it pay to complexify pricing strategies? The answer to that question very much depends on the information set-up. In the complete information case, the answer is basically "no": the difference in revenue between uniform pricing with unconstrained prices and full dynamic pricing is only around 2.0%.[14] This figure sharply contrasts with the 19.3% gain we estimate under incomplete information by comparing Scenarios f.1 and u.2.

Intuitively, dynamic pricing still helps in the complete information case because of the uncertainty on the demand process. But the possibility to adjust the pricing strategy as one learns about (ξ_{aT}, ξ_{bT}) (or, equivalently, (η_{aT}, η_{bT})) in the incomplete information set-up plays a much more important role. To shed light on this point, we decompose the variance of the demand under the optimal uniform pricing in incomplete information into two parts:

$$\mathbb{E}\left[\mathbb{V}(D_{dT}(0,1,p^u_{dT})|W_T)\right] = \mathbb{E}[\mathbb{V}(D_{dT}(0,1,p^u_{dT})|\xi_{dT})] + \mathbb{E}\left[\mathbb{V}(\mathbb{E}[D_{dT}(0,1,p^u_{dT})|\xi_{dT}]|W_T)\right],$$

where p^u_{dT} is the optimal price under uniform pricing for destination d and train T. Even though they both involve $g_0(W_T)$, one can show that the two terms in this de-

[13]These are approximations obtained by dividing each lower bounds and each upper bounds. The exact bounds on the ratios are hard to obtain because the revenues under uniform pricing and constrained prices do not take a simple form. The approximation we use works well on other ratios, for which we can compute the exact bounds.

[14]This empirical finding is consistent with simulation results in operational research and empirical results in economics. For example, Zhao and Zheng (2000) shows a similar improvement by between 2.4% and 7.3%. Williams (2022) estimates a revenue improvement due to optimal dynamic pricing of around 2% in airline industry.

composition are point identified. For intermediate and final destinations respectively, the variation of the demand process (the first term) only explains on average 1.3% and 0.9% of the total variance.

Now, even in the incomplete information set-up, one need not consider complex pricing strategies to obtain revenues close to the optimal ones. First, restricting to stopping-time pricing strategies incurs virtually no loss, compared to "full" dynamic pricing. By changing prices only when a purchase is observed, the firm can secure around 99.8% of the revenue gain from uniform pricing to dynamic pricing regimes (comparing here Scenarios s.5 and f.1). Considering pricing strategies with 12 fare classes, as in reality but with possibly decreasing prices, still yield revenues in between 98.7% and 99.1% of the revenues under full dynamic pricing (comparing here Scenarios s.4 and f.1).

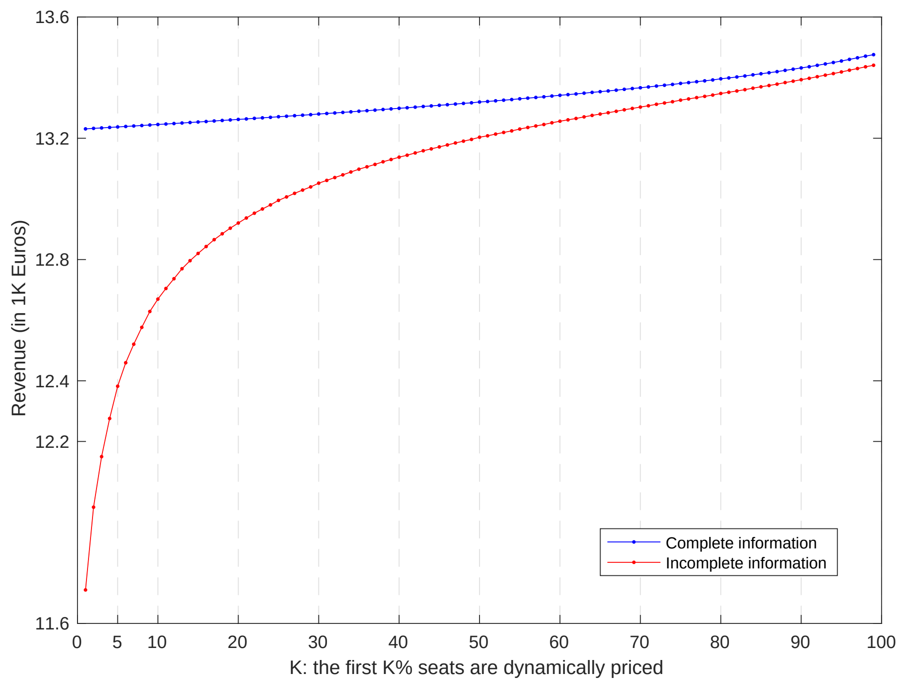
How fast does information accumulate? First, the tiny difference between the gains of full dynamic pricing under complete and incomplete information shows that revenue management is an effective instrument for demand learning. By learning from consumers' purchases in a Bayesian way, it can gradually pin down the uncertainty on the overall demand. Pricing decision then takes this renewed information into account, improving total revenue. And actually, this demand learning can compensate almost all revenue loss due to ex ante uncertainty on demand. The difference in revenue under optimal uniform pricing between incomplete and complete information is around 2K€ (comparing Scenarios u.4 and u.2), while this difference decreases to around 0.03K€ only under optimal dynamic pricing (see Scenarios f.2 and f.1). This finding is in line with Lin (2006), who reports a similar near-optimality of demand learning in a simulation study.

The reason of this very modest loss compared to the complete information set-up is that information accumulates quickly. To illustrate this point, we simulate expected revenues under a class of intermediate stopping-time pricing strategies, where the firm is only allowed to dynamically price the first K% seats, turning to uniform pricing for the remaining seats. Thus, $K = 0$ and $K = 100$ correspond respectively to the optimal uniform and stopping-time pricing strategies.[15] By quantifying the revenue

[15]As the pricing strategies in Theorem 4.1, we show in Appendix D that one can partially iden-

gain from K to $K+1$, we can characterize how much can be marginally gained from being able to extract information on demand from additional purchases (1% of total seats) and optimally adjust its pricing. Figure 1 displays the lower bounds of the optimal revenues under these intermediate pricing strategies under complete (blue) and incomplete (red) information for $K = 1, ..., 100$.[16]

Figure 1: Revenues (lower bound) under intermediate pricing strategies



Under incomplete information, demand learning is rather quick, as we can see from the important concavity of the red line. With just $K = 5$, the firm already achieves a revenue equal to the observed one; by learning from 50% of the seats, it obtains a revenue around 3% lower than that of the complete information. On the other hand, the blue line shows that the revenue gains under complete information are small. The incremental revenue from K to $K+1$ is almost constant and barely reaches 3€. This latter result could be expected, given that the difference between uniform pricing and the full stopping-time pricing is small under complete information. The

tify the optimal revenues with intermediate stopping-time pricing strategies using the procedure described in Section 4.4.

[16]We also simulate the upper bounds of these revenues. The obtained curve is very similar.

striking difference in the pattern of marginal gain between complete and incomplete information settings is also in line with our previous findings: in terms of revenue improvement by dynamic pricing, the effect of learning overall demand (ξ_{aT}, ξ_{dT}) is more remarkable than that of pinning down the uncertainty in demand process when (ξ_{aT}, ξ_{dT}) is fixed.

5.3 Tests and robustness checks

In this section, we first test the plausibility of Assumptions 1 and 3, on which the identification of θ_0 relies. Next, we relax the assumption of a time-invariant price elasticity. Then, we consider alternative parametric specifications. Finally, we explore the effect of specific routes on our results.

Test of Assumptions 1 and 3. These assumptions imply that the proportions $n_{bkT}/(n_{akT}+n_{bkT})$ remain constant through fare classes k satisfying $p_{bkT} = p_{akT}$. A convenient way to check this is to restrict ourselves to two routes, Paris-Marseille and Paris-Mulhouse, for which $p_{bkT} = p_{akT}$ for all $k \in \{1, ..., K\}$. By taking the first fare class as a reference, we simply regress $n_{bkT}/(n_{akT}+n_{bkT})$ on the other 11 fare class dummies and train fixed effects. We then test whether the coefficients of the fare class dummies are equal to zero.

The results are presented in Table 5. As emphasized by the top panel, most coefficients are not significant, despite the large number of observations (453 and 499 for the two routes). For Paris-Marseille, the p-value of the joint test is larger than 0.05. For Paris-Mulhouse, the p-value is lower, but it appears that this result is mostly driven by the last fare classes (the joint test for nullity of the first 10 classes has a p-value of 15%). The coefficients of the last two fare classes are indeed positive and quite large for this route, indicating that there would be more "late purchasers" for Mulhouse than for Strasbourg.

To see whether this pattern could influence our results beyond this specific route, we re-estimate ε using only the first 10 fare classes. We obtain a price elasticity of -4.86, which is thus somewhat higher in absolute value than the baseline estimate of -4.04 obtained with the 12 fare classes. We then recomputed the identified sets of counterfactual revenues for Scenarios u2, u4, s5, s10, f1 and f2 (as they are the simplest to compute). The optimal revenues are slightly higher but with differences never exceeding 3.3% on the lower bounds and 1.2% on the upper bounds.

Table 5: Test of the separability in Assumption 1

	Paris-Marseille	Paris-Mulhouse
Fare class	Coefficient estimates	
2	-0.019	0.003
3	-0.041^{***}	-0.008
4	-0.019	-0.010
5	-0.004	-0.009
6	-0.005	-0.025
7	0.002	-0.020
8	-0.003	0.009
9	0.033	0.026
10	-0.003	0.041
11	-0.03	0.109^{***}
12	-0.025	0.168^{***}
Joint nullity test	p-values	
2-12	0.053	0.0004
Average ratio	0.589	0.249

Notes: Coefficient estimates of the regression of $n_{bkT}/(n_{akT} + n_{bkT})$ on train fixed effects and fare class dummies (fare class 1 being the reference).

Time-varying price elasticities. One could expect that consumers purchasing their tickers earlier would be more price elastic than those buying their tickets late. For instance, the latter could include more business travelers. If so, the assumption of a time-invariant price elasticity would be violated. To test for this condition, we replace ε in (3) by $\varepsilon_{\text{early}}1\{k \leq S\} + \varepsilon_{\text{late}}1\{k > S\}$ for some threshold S that we vary. In other words, we distinguish price elasticity of early purchasers, defined as those who purchase a ticket with price in fare classes inferior or equal to S, from that of late purchasers who pay the price in fare classes superior to S. We then compare

$\varepsilon_{\text{early}}$ to $\varepsilon_{\text{late}}$ to assess the extent to which the assumption of a time-invariant price elasticity holds, and the impact of relaxing this condition on counterfactual revenues.

The results are displayed in Table 6. We consider threshold values S equal to 9, 10 and 11. In the three cases, "early purchasers" are estimated to be more price elastic than "late purchasers", with the estimated price elasticity of the former being greater (in absolute value) than the baseline estimate in Table 3 (-4.04) but still close to its upper bound of the 95% confidence interval.

Table 6: Test of time-invariant elasticity

Segmentation on purchasers			
Early purchasers	Late purchasers	$\varepsilon_{\text{early}}$	$\varepsilon_{\text{late}}$
1-9	10-12	4.76 (0.077)	3.01 (0.11)
1-10	11-12	4.55 (0.074)	3.17 (0.14)
1-11	12	4.53 (0.070)	2.88 (0.22)

Notes: In the first line, early purchasers are supposed to buy their tickets before the ninth fare class closes. Analytical standard errors are under parentheses.

Next, to assess the robustness of the time-invariance elasticity condition affects counterfactual revenues, we simulate some scenarios by explicitly considering early and late purchasers (with $\varepsilon_{\text{early}}$ and $\varepsilon_{\text{late}}$, respectively) in the demand model. Specifically, we set $S = 10$, re-estimate the demand using the procedure described in Sections 4.2 and 4.3 with $\varepsilon_{\text{early}}$ and $\varepsilon_{\text{late}}$, and compute the revenue with the optimal uniform pricing under complete information (scenario u.4 in Table 4). We find that the estimates of destination-train-specific effect, β_0, and the parameters of the Gamma distribution, λ_0, are close to the baseline results. Furthermore, the simulated revenue is close to that in the scenario u.4 in Table 4. Both findings suggest that despite the difference in price elasticity of early and late purchasers, the results are robust to the assumption of a time-invariant price elasticity. We refer to the Online Appendix E for more details on the estimation method and the results.

Alternative parametric specifications. We conduct two robustness checks. First, we simulate counterfactual revenues with a lognormal specification on η_{dT} instead of a gamma distribution (Assumption 2(ii)). The drawback of a lognormal specification is that it is not conjugate with the Poisson distribution. Then, the updated distribution of η_{dT} in the incomplete information set-up takes a complicated form, making it very difficult to compute counterfactual scenarios. Nevertheless, this issue does not appear for uniform pricing and complete information. Table 7 shows the results for Scenarios u.2, u.4, s.10 and f.2. Even if the bounds are wider than in the baseline specification, the results are similar.

Table 7: Counterfactual revenues with log-normally distributed η_{dT}

Scenario	Point or set estimate (in K€)	
Observed pricing strategy	12.21	
Uniform pricing strategy	Log-normal specification	Baseline
u.2 Incomplete inf., uncons. prices	[10.03, 12.30]	[11.29, 12.31]
u.4 Complete inf., uncons. prices	[12.67, 15.36]	[13.23, 14.42]
Complete information		
s.10 Stopping-time pricing strategy	[12.90, 15.66]	[13.48, 14.70]
f.2 "Full" dynamic pricing strategy	[12.92, 15.68]	[13.50, 14.72]

Notes: The baseline results correspond to those in Table 4. See the notes of that table for more details.

Second, we have focused so far on the demand model corresponding to Column I in Table 3. We did so because counterfactual revenues are harder to compute under the richer specification corresponding to Column II in the same table. Nevertheless, we were able to compute counterfactual revenues with this specification for a few scenarios. The results, presented in Table 8, are hardly affected.

Effect of specific routes. Table 2 show that the routes to Marseille and Côte basque have unusually high and low loads, so one may worry that revenue management was very different for these lines. We resimulate the counterfactual revenues by excluding these two routes. The results are hardly affected, with changes in the bounds by at most 1% over all scenarios.

Table 8: Counterfactual revenues based on Column II in Table 3

Scenario	Point or set estimate (in K€)	
Observed pricing strategy	12.21	
Uniform pricing strategy	Specification II	Baseline
u.1 incomplete information, constrained prices	[11.51, 12.21]	[11.15, 12.21]
u.2 incomplete information, unconstrained prices	[11.77, 12.37]	[11.29, 12.31]
u.3 complete information, constrained prices	[12.85, 13.73]	[12.56, 13.86]
u.4 complete information, unconstrained prices	[13.63, 14.35]	[13.23, 14.42]
Stopping-time pricing strategy		
s.5 incomplete information	[13.83, 14.57]	[13.44, 14.66]
s.10 complete information	[13.87, 14.60]	[13.48, 14.70]
"Full" dynamic pricing strategy		
f.1 incomplete information	[13.86, 14.59]	[13.47, 14.68]
f.2 complete information	[13.89, 14.62]	[13.50, 14.72]

Notes: The baseline results are those of Table 4. See the notes of that table for more details.

6 Conclusion

Though the framework we have developed is taylored to our application, several of our results could be applied to other set-ups. The insight that many counterfactual revenues only depend on price elasticity and total demand, and not on the precise timing of consumers' arrival, is convenient when no details on the dates of purchases are available. Similarly, the censoring issue and the absence of exogenous variations in prices may often occur. Our identification strategy, combining exogenous variations in relative prices and moment inequalities based on basic rationality on consumer's side and weak optimality conditions on the firm's pricing strategy, could then be applied in such contexts. Our results also suggest that such moment inequalities may be quite informative in practice.

References

Aviv, Y. and Pazgal, A. (2002), Pricing of short life-cycle products through active learning. Working paper.

Bitran, G. R. and Mondschein, S. V. (1997), ‘Periodic pricing of seasonal products in retailing’, *Management science* **43**(1), 64–79.

Boyd, E. A. and Bilegan, I. C. (2003), ‘Revenue management and e-commerce’, *Management science* **49**(10), 1363–1386.

Brémaud, P. (1981), *Point processes and queues: martingale dynamics*, Vol. 50, Springer.

Brumelle, S. and McGill, J. (1993), ‘Airline seat allocation with multiple nested fare classes’, *Operations Research* **41**(1), 127–137.

Cho, S., Lee, G., Rust, J. and Yu, M. (2018), Optimal dynamic hotel pricing. Working paper.

Cho, S., Lee, G., Rust, J. and Yu, M. (2019), Semi-parametric instrument-free demand estimation: relaxing optimality and equilibrium assumptions. Working paper.

den Boer, A. V. (2015), ‘Dynamic pricing and learning: historical origins, current research, and new directions’, *Surveys in operations research and management science* **20**(1), 1–18.

den Boer, A. V. and Zwart, B. (2015), ‘Dynamic pricing and learning with finite inventories’, *Operations research* **63**(4), 965–978.

D’Haultfœuille, X. and Rathelot, R. (2017), ‘Measuring segregation on small units: A partial identification analysis’, *Quantitative Economics* **8**(1), 39–73.

Feng, Y. and Gallego, G. (1995), ‘Optimal starting times for end-of-season sales and optimal stopping times for promotional fares’, *Management Science* **41**(8), 1371–1391.

Feng, Y. and Xiao, B. (2000), ‘Optimal policies of yield management with multiple predetermined prices’, *Operations Research* **48**(2), 332–343.

Gallego, G. and Van Ryzin, G. (1994), ‘Optimal dynamic pricing of inventories with stochastic demand over finite horizons’, *Management science* **40**(8), 999–1020.

Huang, Y., Ellickson, P. B. and Lovett, M. J. (2022), ‘Learning to set prices’, *Journal of Marketing Research* **59**, 411–434.

Jevons, D., Meaney, A., Robins, N., Dargay, J., Preston, J., Goodwin, P. and Wardman, M. (2005), ‘How do passengers respond to change?’, Working paper, Transport Studies Unit and Oxford University Centre for the Environment.

Lazarev, J. (2013), The welfare effects of intertemporal price discrimination: An empirical analysis of airline pricing in U.S. monopoly markets. Working paper.

Lee, A. (1990), Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process, PhD thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Lin, K. Y. (2006), ‘Dynamic pricing with real-time demand learning’, *European Journal of Operational Research* **174**(1), 522–538.

Littlewood, K. (1972), ‘Forecasting and control of passenger bookings’, *Airline Group International Federation of Operational Research Societies Proceedings, 1972* **12**, 95–117.

McAfee, R. P. and te Velde, V. (2008), ‘Dynamic pricing with constant demand elasticity’, *Production and Operations Management* **17**(4), 432–438.

Phillips, R. L. (2021), *Pricing and revenue optimization*, Stanford university press.

Sauvant, J. (2002), ‘Le transport ferroviaire de voyageurs en france : enfin un bien normal?’, *Note de synthèse du SES* .

Stefanescu, C. (2012), ‘Multivariate demand: Modeling and estimation from censored sales’, Working paper.

Swan, W. (1990), ‘Revenue management forecasting biases’, Working paper, Boeing Commercial Aircraft, Seattle.

Talluri, K. and van Ryzin, G. (2005), *The theory and practice of revenue management*, Springer Verlag.

Wardman, M. (1997), 'Inter-urban rail demand elasticities and competition in Great Britain: Evidence from direct demand models', *Transportation Research* **33E**(1), 15–28.

Wardman, M. (2006), 'Demand for rail travel and the effects of external factors', *Transportation Research* **42E**, 129–148.

Wardman, M., Lythgoe, W. and Whelan, G. (2007), 'Rail passenger demand forecasting: cross-sectional models revisited', *Research in Transportation Economics* **20**(3), 119–152.

Williams, K. R. (2022), 'The welfare effects of dynamic pricing: Evidence from airline markets', *Econometrica* **90**(2), 831–858.

Zhao, W. and Zheng, Y.-S. (2000), 'Optimal dynamic pricing for perishable assets with nonhomogeneous demand', *Management science* **46**(3), 375–388.

A Proofs

A.1 Theorem 4.1

The counterfactual revenues depend at most on the parameters of the model, which are ε, the distribution of (ξ_{aT}, ξ_{bT}) and $b_T(.)$. Thus, to prove the result, it suffices to show that these counterfactual revenues only depend on ε and the distribution of (ξ_{aT}, ξ_{bT}).

To this end, fix $I \in \{c, i\}$ and $r \in \{u, f, s, sM, sM+\}$ and let $B_T(t) = \int_0^t b_T(u)du$. By assumption, $B_T(.)$ is continuous and strictly increasing. Thus, it admits an inverse that is defined on $[0, 1]$. Then, for $d \in \{a, b\}$, $t \in [0, 1]$, $A \subset [0, \infty)$ and $p > 0$, let us define $\widetilde{V}_{dT}$ as

$$\widetilde{V}_{dT}([0, t], A) = V_{dT}([0, B_T^{-1}(t)], A).$$

Then, by Assumption 1, $\widetilde{V}_{dT}$ is a Poisson process with intensity $(t, p) \mapsto \xi_{dT}\varepsilon p^{-\varepsilon-1}$. Let us denote by $\widetilde{R}_r^I$ the optimal revenue associated with $(\widetilde{V}_{aT}, \widetilde{V}_{bT})$. We prove below that $\widetilde{R}_r^I = R_r^I$. This proves the result because the distribution of $\widetilde{V}_{dT}$ only depends on ε and the distribution of (ξ_{aT}, ξ_{bT}).

Let us consider the pricing strategy $(p_{r,a}^I(.), p_{r,b}^I(.))$ associated with the processes (V_{aT}, V_{bT}), satisfying the constraints associated with r and I and leading (on expectation) to the optimal revenue R_r^I.[17] Note in particular that $p_{r,a}^I(t)$ should only depend on purchases up to t and $b_T(.)$ if $I = c$. Now, let us define, for $d \in \{a, b\}$ and $t \in [0, 1]$,

$$\widetilde{p}_{r,d}^I(t) = p_{r,d}^I \left[B_T^{-1}(t)\right].$$

By construction, $\widetilde{p}_{r,d}^I(t)$ only depends on purchases (corresponding to $\widetilde{V}_{dT}$) up to t and B_T if $I = c$. Also, as $p_{r,d}^I$, it satisfies all the constraints associated with r, since no constraints are related to time. Hence, $(\widetilde{p}_{r,a}^I, \widetilde{p}_{r,b}^I)$ is a feasible pricing strategy. As such, it leads to an expected revenue that is lower than $\widetilde{R}_r^I$. Also, by construction, we obain with $(\widetilde{p}_{r,a}^I, \widetilde{p}_{r,b}^I)$, associated to $(\widetilde{V}_{aT}, \widetilde{V}_{bT})$, the same purchases at the same prices as with the pricing strategy $(p_{r,a}^I, p_{r,b}^I)$ when associated to (V_{aT}, V_{bT}). Thus, at

[17] There may be no such pricing strategy, but only sequences of pricing strategies with corresponding expected revenue tending to R^I. If so, we just replace $p_r^I(.)$ by the corresponding sequence in the rest of the proof.

the end, $R_r^I \leq \widetilde{R}_r^I$. By the same reasoning, just switching the role of V_{dT} and $\widetilde{V}_{dT}$, we obtain $\widetilde{R}_r^I \leq R_r^I$. Hence, $R_r^I = \widetilde{R}_r^I$ and the result follows.

A.2 Theorem 4.2

Hereafter, ℓ denotes the dimension of X_{dT}, Supp(A) denotes the support of any random variable A. and $\Delta X_T := X_{bT} - X_{aT}$. We first state Assumption 5 appearing in the statement of the theorem.

Assumption 5 *There exists a component X_{djT} of X_{dT} such that Supp$(\Delta X_{jT}|\Delta X_{-jT}) = \mathbb{R}$, where X_{d-jT} is the vector stacking all the components of X_{dT} except X_{djT}. Also, there exists $x_{-j}^1, x_{-j}^2, ..., x_{-j}^\ell \in$ Supp$(X_{b-jT} - X_{a-jT})$ such that the matrix*

$$M := \begin{pmatrix} {x_{-j}^1}' - {x_{-j}^2}' \\ \vdots \\ {x_{-j}^1}' - {x_{-j}^\ell}' \end{pmatrix}$$

is nonsingular. Finally, $\sup \textit{Supp}\left(\sum_{k=1}^K n_{akT} + n_{bkT}\right) \geq K + 1$.

The most restrictive condition is $\text{Supp}(\Delta X_{jT}|\Delta X_{-jT}) = \mathbb{R}$. However, the proof below reveals that identification is not achieved at infinity. The large support condition simply ensures that we can produce the compensating variations used in the proof. Also, the nonparametric identification of the distribution of η_{bT}/η_{aT} can be obtained without large support, using the analyticity of the density of the logistic distribution. The last condition ($\sup \text{Supp}\left(\sum_{k=1}^K n_{akT} + n_{bkT}\right) \geq K + 1$) easily holds in our application. It shows that our result does not require a large number of units of the perishable good (namely, tickets for a given train in our application) to be applicable.

1. First, note that because the realization of τ_k is determined by the Poisson process before τ_k and is independent of $D_{dT}(\tau_k, \tau_{k+1}, ; p_{dkT})$ for $d \in \{a, b\}$, it suffices to show (3) if τ_k is replaced by any fixed number that we suppose equal to 0 without loss of generality. To ease the exposition, we often omit the index T hereafter and define $\mu_d = \xi_d p_{dk}^{-\varepsilon}$ and $\rho = \mu_a/(\mu_a + \mu_b)$. We also introduce $D_{d,\tau_n} = D_{dT}(0, \tau_n; p_{dkT})$ for $d \in \{a, b\}$, $D_{\tau_n} = D_{a,\tau_n} + D_{b,\tau_n}$ and $\tau_n = \inf\{t > 0 : D_t \geq n\} \wedge 1$. We will show that for all $n \geq 1$,

$$D_{a,\tau_n}|D_{\tau_n}, b_T(.), \xi_a, \xi_b \sim \text{Binomial}\left(D_{\tau_n}, \rho\right). \tag{11}$$

Given the previous discussion and because the right-hand side of (11) does not depend on $b_T(.)$, (3) will follow from (11).

To prove (11), we introduce, for any $n \geq 1$, the hitting times $\sigma_n = \inf\{t \in [0,1] : D_t \geq n\}$, with $\sigma_n = 2$ if $D_1 < n$. Let us also fix $\underline{t} \in (0,1)$ and let us partition the interval $I = [\underline{t}, 1]$ into m intervals $I_1, ..., I_m$ of equal length $\Delta t = (1-\underline{t})/m$. Finally, for all $c \leq n$, let

$$q_{c,n;k} = \Pr[D_{a,\sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I_k]. \tag{12}$$

By Lemma C.1, there exists (c_l, c_r), independent of k and m, such that for all $k = 1, ..., m$,

$$-c_l(1+n)\Delta t \leq q_{c,n;k} - \binom{n}{c}\rho^c(1-\rho)^{n-c} \leq c_r \Delta t.$$

Moreover, we have

$$\begin{aligned}\Pr[D_{a,\sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I] &= \frac{\sum_{k=1}^m \Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k]}{\sum_{k=1}^m \Pr[D_{\sigma_n} = n, \sigma_n \in I_k]} \\ &\in \left[\min_{k=1,...,m} q_{c,n;k}, \max_{k=1,...,m} q_{c,n;k}\right].\end{aligned}$$

Consequently,

$$-c_l(1+n)\Delta t \leq \Pr[D_{a,\sigma_n} = c | D_{\sigma_n} = n, \sigma_n \in I] - \binom{n}{c}\rho^c(1-\rho)^{n-c} \leq c_r \Delta t.$$

By letting $m \to \infty$ and then let $\underline{t} \to 0$, we obtain

$$\Pr[D_{a,\sigma_n} = c | D_{\sigma_n} = n, \sigma_n \leq 1] = \binom{n}{c}\rho^c(1-\rho)^{n-c}. \tag{13}$$

Now, because $D_{\tau_n} = n$ if and only if $\sigma_n \leq 1$, we obtain (11) in this case. Further, because $D_{\tau_n} = n' < n$ if and only if $D_1 = n'$ and $\sigma_n = 2$, we have

$$\begin{aligned}\Pr[D_{a,\tau_n} = c | D_{\tau_n} = n'] &= \Pr[D_{a,1} = c | D_1 = n', \sigma_n = 2] \\ &= \Pr[D_{a,1} = c | D_1 = n'] \\ &= \binom{n'}{c}\rho^c(1-\rho)^{n'-c}.\end{aligned}$$

Thus, (11) also holds when $D_{\tau_n} = n'$, $n' < n$. The result follows.

2. Consider two fare classes k, k' such that $p_{ak}/p_{bk} \neq p_{ak'}/p_{bk'}$ (hereafter, we implicitly reason conditional on prices). Fix $x \in \mathbb{R}$ and let

$$\widetilde{x} = x - \frac{\varepsilon}{\beta_{0j}} \ln\left(\frac{p_{ak}}{p_{bk}}\frac{p_{bk'}}{p_{ak'}}\right). \tag{14}$$

Then, $x\beta_{0j} - \varepsilon\ln(p_{ak}/p_{bk}) = \widetilde{x}\beta_{0j} - \varepsilon\ln(p_{ak'}/p_{bk'})$. In turn, given the index structure,

$$\begin{aligned}&\Pr(n_{bkT} = n_b | n_{akT} + n_{bkT} = n, \Delta X_{-jT}, \Delta X_{jT} = x)\\ =&\Pr(n_{bk'T} = n_b | n_{ak'T} + n_{bk'T} = n, \Delta X_{-jT}, \Delta X_{jT} = \widetilde{x}).\end{aligned}$$

Conversely, there is a single solution $\widetilde{x}$ to this equation, given by (14). Hence, $\widetilde{x}$ and thus β_{0j} are identified. Similarly, for any two $x_{-j} \neq \widetilde{x}_{-j}$ in the support of ΔX_{-jT},

$$\begin{aligned}&\Pr(n_{akT} = n_b | n_{akT} + n_{bkT} = n, \Delta X_{-jT} = x_{-j}, \Delta X_{jT} = x)\\ =&\Pr(n_{akT} = n_b | n_{akT} + n_{bkT} = n, \Delta X_{-jT} = \widetilde{x}_{-j}, \Delta X_{jT} = \widetilde{x}).\end{aligned}$$

if and only if $x\beta_{0j} + x'_{-j}\beta_{0-j} = \widetilde{x}\beta_{0j} + \widetilde{x}'_{-j}\beta_{0-j}$. By considering $x^1_{-j}, ..., x^\ell_{-j}$ as in Assumption 5, we obtain $M\beta_{0j} = y$ for some identified vector y. Since M is nonsingular, β_{0-j} is identified.

Finally, we show the nonparametric identification of the cumulative distribution function (cdf) F of $\ln(\eta_{aT}/\eta_{bT})$. Since $\sup\text{Supp}\left(\sum_{k=1}^K n_{akT} + n_{bkT}\right) \geq K+1$, there exists a fare class k for which $\sup\text{Supp}(n_{akT} + n_{bkT}) \geq 2$. Fix $n \geq 2$. The distribution of $n_{bkT}|n_{akT} + n_{bkT} = n, \Delta X_T = x$ is a binomial mixture, with mixture distribution G_x, say. Then (see, e.g. D'Haultfœuille and Rathelot, 2017), the first n moments of G_x are identified. In particular, we identify $\int_0^1 p(1-p)dG_x(p)$. Now, given the structure of the problem,

$$\int_0^1 p(1-p)dG_x(p) = \int \Lambda'(x'\beta_0 - u)dF(u),$$

with $\Lambda' = \Lambda(1-\Lambda)$ the density of the logistic distribution. By varying x_j over $\mathbb{R}$, we thus identify the distribution of $U+V$, where U and V are independent, U is logistic and $V \sim F$. Taking the Fourier transform, we thus identify $\Psi_U \times \Psi_V$, where Ψ_U and Ψ_V (resp. Ψ_V) denotes the characteristic function of U (resp. V). Since $\Psi_U(t) = \pi t/\sinh(\pi t) \neq 0$, Ψ_V is identified. Hence, F is identified as well.

B Demand Estimation with Aggregated Data

The difference between our results and those from studies relying on aggregated data comes precisely from the fact that we dispose of micro-level data. The approach based on aggregate data is likely to to bias upwards the price-elasticity estimates. Average prices are endogenous, since the weights associated to each price or, equivalently, to each fare class, is fully driven by the demand. Basically, trains in high demand are likely to have a few number of seats available at a low price, resulting in a higher average price. To illustrate this point, we aggregate our micro data and estimate the corresponding price elasticities. For instance, we propose to aggregate data over fare classes at the train level, and thus to consider an average price for every train. Then we regress the logarithm of total purchases on the logarithm of this average price.

We first aggregate the data at the train and destination level. Let Q_{dT} be the total quantity of tickets purchased for destination d in the train T, $Q_{dT} = \sum_{k=1}^{K} n_{dkT}$. The corresponding average price $\overline{p}_{dT}$ is given by:

$$\overline{p}_{dT} = \frac{\sum_{k=1}^{K} n_{dkT} p_{dkT}}{\sum_{k=1}^{K} n_{dkT}}.$$

We then consider a constant elasticity demand model with train fixed effects:

$$\ln(Q_{dT}) = -\epsilon \ \ln(\overline{p}_T) + \delta_T + \xi_d + \nu_{dT}, \tag{15}$$

where ξ_d accounts for a destination-specific component.

We then aggregate further our data at the train level, by considering $Q_T = Q_{aT} + Q_{bT}$ and defining the corresponding average price:

$$\overline{p}_T = \frac{\sum_{d\in\{a,b\}} \sum_{k=1}^{K} n_{dk} p_{dkT}}{\sum_{d\in\{a,b\}} \sum_{k=1}^{K} n_{dk}}.$$

We consider a similar model as (15), except that at that level of aggregation, we cannot include train and destination fixed effects. Instead, we include day of departure and route fixed effects:

$$\ln(Q_T) = -\epsilon \ln(\overline{p}_T) + \delta_{t(T)} + \xi_{r(T)} + \nu_T, \tag{16}$$

where $t(T)$ and $r(T)$ denote the day of departure and the route of train T. Finally, the most aggregated approach consists in aggregating these demands at a weekly or monthly level, either by train route or at the national level.

Results are given in Table 9. The first line presents the price elasticity estimate for the less disaggregated specification. Strikingly, the estimate (-1.02) is already much higher than ours. It is close to the estimate of -0.70 obtained by Sauvant (2002) on SNCF aggregated data. By aggregating further at the train level, we exacerbate the bias and obtain already a positive coefficient (0.15). Aggregating further at the week or at the month level increases further the coefficient, up to 1.14. Using data aggregated at the national level leads to somewhat lower coefficients, but still positive ones (0.14 and 0.56 for weekly and monthly data, respectively).

Table 9: Estimated price elasticities with aggregated data

Model	Price elasticity
Train and destination level (Equation (15))	-1.02 (0.24)
Train level (Equation (16))	0.15 (0.03)
Week× route level	0.29 (0.12)
Month× route level	1.14 (0.40)
Week level (France)	0.14 (0.09)
Month (France)	0.56 (0.33)

Notes: We refer to the text for a detailed explanation of each model.

Online Appendix

C A key lemma for Theorem 4.2

The proof of Point 1 of Theorem 4.2 crucially relies on the following lemma, which we prove below (the notation we use are introduced in the proof of Theorem 4.2).

Lemma C.1 *Suppose that Assumption 1 holds. Then, there exists c_l and c_r, independent of k and m, such that for all $k = 1, ..., m$,*

$$-c_l(1+n)\Delta t \leq q_{c,n;k} - \binom{n}{c}\rho^c(1-\rho)^{n-c} \leq c_r\Delta t. \tag{17}$$

Proof: First, observe that $\{\sigma_n \in I_k\} = \{D_{\underline{t}+(k-1)\Delta t} < n,\ D_{\underline{t}+k\Delta t} \geq n\}$. Then

$$\begin{aligned}
&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k] \\
=&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n,\ D_{\underline{t}+k\Delta t} \geq n] \\
=&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1,\ D_{\underline{t}+k\Delta t} \geq n] \\
&+\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n-1,\ D_{\underline{t}+k\Delta t} \geq n]
\end{aligned} \tag{18}$$

We first show that the second term in (18) is negligible, as being of order $(\Delta t)^2$. Simple algebra shows that if $U \sim \mathcal{P}(\mu)$, then $\Pr(U \geq 2) \leq \mu^2$. Hence,

$$\begin{aligned}
&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} < n-1,\ D_{\underline{t}+k\Delta t} \geq n] \\
\leq&\Pr[D_{\underline{t}+k\Delta t} - D_{\underline{t}+(k-1)\Delta t} \geq 2] \\
\leq&\left[(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds\right]^2 \\
\leq&\left[(\mu_a+\mu_b)\bar{b}\Delta t\right]^2,
\end{aligned}$$

where $\bar{b} = \sup_{t\in[0,1]} b(t)$. Now, the first term in (18) satisfies:

$$\begin{aligned}
&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1,\ D_{\underline{t}+k\Delta t} \geq n] \\
=&\Pr[D_{a,\sigma_n} = n, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1,\ D_{\underline{t}+k\Delta t} = n] \\
+&\Pr[D_{a,\sigma_n} = c, D_{\sigma_n} = n, D_{\underline{t}+(k-1)\Delta t} = n-1,\ D_{\underline{t}+k\Delta t} > n],
\end{aligned}$$

where the second term can be similarly controlled as above:

$$\begin{aligned}\Pr[D_{a,\sigma_n}=c, D_{\sigma_n}=n, D_{\underline{t}+(k-1)\Delta t}=n-1,\ D_{\underline{t}+k\Delta t}>n] &\le \Pr[D_{\underline{t}+k\Delta t}-D_{\underline{t}+(k-1)\Delta t}\ge 2]\\ &\le \left[(\mu_a+\mu_b)\bar{b}\Delta t\right]^2.\end{aligned}$$

As a consequence,

$$\begin{aligned}&\Pr[D_{a,\sigma_n}=c, D_{\sigma_n}=n, D_{\underline{t}+(k-1)\Delta t}=n-1,\ D_{\underline{t}+k\Delta t}=n]\\ <&\Pr[D_{a,\sigma_n}=c, D_{\sigma_n}=n, \sigma_n\in I_k]\\ \le&\Pr[D_{a,\sigma_n}=c, D_{\sigma_n}=n, D_{\underline{t}+(k-1)\Delta t}=n-1,\ D_{\underline{t}+k\Delta t}=n]\\ &+2[(\mu_a+\mu_b)\bar{b}]^2(\Delta t)^2.\end{aligned}\tag{19}$$

Now, we have

$$\begin{aligned}&\Pr[D_{a,\sigma_n}=c, D_{\sigma_n}=n, D_{\underline{t}+(k-1)\Delta t}=n-1,\ D_{\underline{t}+k\Delta t}=n]\\ =&\Pr[D_{a,\sigma_n}=c, D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1]\\ &\Pr[D_{a,\sigma_n}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1]\\ =&\Pr[D_{a,\sigma_n}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c-1]\\ &+\Pr[D_{a,\sigma_n}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c]\\ =&\Pr[D_{a,\underline{t}+k\Delta t}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c-1]\\ &+\Pr[D_{a,\underline{t}+k\Delta t}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c].\end{aligned}\tag{20}$$

Now, by independence between $(D_{a,t})_{t\ge 0}$ and $(D_{b,t})_{t\ge 0}$, and independence between $D_{d,t+s}-D_{d,t}$ and $D_{d,t}$ for all $s>0$ and $d\in\{a,b\}$,

$$\begin{aligned}&\Pr[D_{a,\underline{t}+k\Delta t}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c-1]\\ =&\Pr[D_{a,\underline{t}+(k-1)\Delta t}=c-1]\Pr[D_{a,\underline{t}+k\Delta t}=c|D_{a,\underline{t}+(k-1)\Delta t}=c-1]\\ &\times\Pr[D_{b,\underline{t}+(k-1)\Delta t}=n-c]\Pr[D_{b,\underline{t}+k\Delta t}=n-c|D_{b,\underline{t}+(k-1)\Delta t}=n-c]\\ =&\frac{\mu_a^c\mu_b^{n-c}\left(\int_0^{\underline{t}+(k-1)\Delta t}b_s ds\right)^{n-1}}{(n-c)!(c-1)!}\exp\left\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t}b_s ds\right\}\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t}b_s ds.\end{aligned}$$

Similarly,

$$\begin{aligned}&\Pr[D_{a,\underline{t}+k\Delta t}=c,\ D_{\underline{t}+k\Delta t}=n,\ D_{\underline{t}+(k-1)\Delta t}=n-1, D_{a,\underline{t}+(k-1)\Delta t}=c]\\ =&\frac{\mu_a^c\mu_b^{n-c}\left(\int_0^{\underline{t}+(k-1)\Delta t}b_s ds\right)^{n-1}}{(n-c-1)!c!}\exp\left\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t}b_s ds\right\}\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t}b_s ds.\end{aligned}$$

By plugging the last two equalities into (20), we obtain

$$
\begin{aligned}
&\Pr[D_{a,\sigma_n} = c,\ D_{\underline{t}+k\Delta t} = n,\ D_{\underline{t}+(k-1)\Delta t} = n-1] \\
=&\frac{n\mu_a^c\mu_b^{n-c}\left(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds\right)^{n-1}}{(n-c)!c!} \exp\left\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\right\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds.
\end{aligned}
$$

Inequality (19) then becomes

$$
\begin{aligned}
&\frac{n\mu_a^c\mu_b^{n-c}(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{(n-c)!c!} \exp\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
<&P(D_{a,\sigma_n} = c, D_{\sigma_n} = n, \sigma_n \in I_k) \\
\leq&\frac{n\mu_a^c\mu_b^{n-c}(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{(n-c)!c!} \exp\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
&+2[(\mu_a+\mu_b)\bar{b}]^2(\Delta t)^2. \qquad (21)
\end{aligned}
$$

By summing (21) over $c = 0, 1, ..., n$, we obtain

$$
\begin{aligned}
&\frac{n(\mu_a+\mu_b)^n(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds \\
<&P(D_{\sigma_n} = n, \sigma_n \in I_k) \\
\leq&\frac{n(\mu_a+\mu_b)^n(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds. \\
&+2(n+1)[(\mu_a+\mu_b)\bar{b}]^2(\Delta t)^2. \qquad (22)
\end{aligned}
$$

By combining (21), (22), and (12), we obtain the following inequalities:

$$
-c_{l,k}(1+n)(\Delta t)^2 \leq q_{c,n;k} - \binom{n}{c}\rho^c(1-\rho)^{n-c} \leq c_{r,k}(\Delta t)^2,
$$

where

$$
c_{r,k} = \frac{2(n+1)[(\mu_a+\mu_b)\bar{b}]^2}{\frac{n(\mu_a+\mu_b)^n(\int_0^{\underline{t}+(k-1)\Delta t} b_s ds)^{n-1}}{n!} \exp\{-(\mu_a+\mu_b)\int_0^{\underline{t}+k\Delta t} b_s ds\}(\mu_a+\mu_b)\int_{\underline{t}+(k-1)\Delta t}^{\underline{t}+k\Delta t} b_s ds},
$$

$$
c_{l,k} = c_{r,k}\binom{n}{c}\rho^c(1-\rho)^{n-c}.
$$

Finally, note that $c_{r,k}\Delta t \leq c_r$ where

$$
c_r = \frac{2(n+1)[(\mu_a+\mu_b)\bar{b}]^2 \exp\{(\mu_a+\mu_b)\int_0^1 b_s ds\}}{\frac{n(\mu_a+\mu_b)^n(\int_0^{\underline{t}} b_s ds)^{n-1}}{n!}(\mu_a+\mu_b)\inf_{s\in I} b_s}.
$$

Moreover, c_r does not depend on k and m. Finally, defining $c_l = c_r\binom{n}{c}\rho^c(1-\rho)^{n-c}$, c_l does not depend on k and m either, and (17) holds for all $k = 1, ..., m$.

D Expressions for the counterfactual revenues

In this appendix, we list and prove the formulas for the counterfactual revenues. The formulas are given conditional on (X_{aT}, X_{bT}, W_T) and for simplicity, we assume here that C_{aT} and C_{bT} are constant; if not, the results should just be seen conditional on (C_{aT}, C_{bT}). Due to the independence between η_{aT} and η_{bT} in Assumption 2(i) and the pre-allocation of capacity, we can separably simulate the counterfactual revenue for each destination, and sum them up to obtain the revenue for train T. Consequently, we focus on pricing for destination d served by train T to simplify the exposition.

We both consider arbitrary distributions for ξ_{dT} and the gamma distribution in Assumption 2. Hereafter, in addition to the scenarios described in Section 3.3, we consider intermediate pricing strategies described in Section 5.2, see Figure 1 and the discussion above it. The corresponding revenues are denoted by R^I_{iK}, with $I \in \{c, i\}$ and where $K \in [0, 100]$ indexes the proportion of seats that are dynamically priced.

Under Assumption 2(i), revenue formulas under both complete and incomplete information have the following form: for $I \in \{c, i\}$ and $r \in \{u, f, s, sM, sM+, iK\}$,

$$R^I_r = \alpha^I_r(\varepsilon, C_{dT}, f) \exp\{X'_{dT}\beta_0/\varepsilon\} g_0^{1/\varepsilon}(W_T),$$

where $f(.)$ denotes the distribution of η_{dT}. In each scenario, we will display the result for a general $f(.)$, and then apply it to the case where f is the density of a gamma distribution $\Gamma(\lambda_{d0}, 1)$. To simplify the formulas and the proofs, we only specify $\alpha^I_r(\varepsilon, C_{dT}, f)$ in each case. Finally, $D(q)$ denotes hereafter a random variable satisfying $D(q) \sim \mathcal{P}(q)$.

D.1 Complete information

D.1.1 Formulas

Uniform pricing $\alpha^c_u = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge C_{dT}] \right\} \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$.

Full-dynamic pricing $\alpha^c_f = \alpha^c_{C_{dT},f} \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$, where $\alpha^c_{0,f} = 0$ and for all $k \geq 1$, $\alpha^c_{k,f} = (\alpha^c_{k,f} - \alpha^c_{k-1,f})^{1-\varepsilon} (1 - 1/\varepsilon)^{\varepsilon-1}$.

Stopping-time pricing $\alpha_s^c = \alpha_{C_{dT},s}^c \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$, where $\alpha_{0,s}^c = 0$ and for all $k \geq 1$,

$$\alpha_{k,s}^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1-e^{-q}) + \alpha_{k-1,s}^c \int_0^1 qe^{-sq}(1-s)^{\frac{1}{\varepsilon}} ds \right\}.$$

Stopping-time pricing with M fares $R_{sM}^c = \alpha_{C_{dT},sM}^c \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$, where $\alpha_{C_{dT},sM}^c = \max_{q>0} \alpha_{C_{dT},M}(q)$, $\alpha_{k,0}(q) = q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge k]$ and for all $k \in \{1,..,C_{dT}\}$,

$$\alpha_{k,m}(q) = \max \Bigg\{ \int_0^1 qe^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1,m\wedge(k-1)}(q(1-z))(1-z)^{\frac{1}{\varepsilon}}\right] dz,$$
$$\max_{q>0} \int_0^1 qe^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1,m-1}(q(1-z))(1-z)^{\frac{1}{\varepsilon}}\right] dz \Bigg\}.$$

Stopping-time pricing with M increasing fares $\alpha_{sM+}^c = \alpha_{C_{dT},sM+}^c \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$, where $\alpha_{C_{dT},sM+}^c = \max_{q>0} \alpha_{C_{dT},M}^+(q)$ with $\alpha_{k,0}^+(q) = \alpha_{k,0}(q)$ and

$$\alpha_{k,m}^+(q) = \max \Bigg\{ q \int_0^1 e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1,m\wedge(k-1)}^+(q(1-z))(1-z)^{\frac{1}{\varepsilon}}\right] dz,$$
$$\max_{q' \in (0,q]} q' \int_0^1 e^{-q'z} \left[q'^{-\frac{1}{\varepsilon}} + \alpha_{k-1,m-1}^+(q'(1-z))(1-z)^{\frac{1}{\varepsilon}}\right] dz. \Bigg\}.$$

Intermediate-K stopping-time pricing $\alpha_{iK}^c = \alpha_{C_{dT},iK}^c \mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right]$, where $\alpha_{C_{dT}(1-K\%),iK}^c = \max_{q>0} q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge (C_{dT}(1-K\%))]$ and for $k > C(1-K\%)$,

$$\alpha_{k,iK}^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1-e^{-q}) + \alpha_{k-1,iK}^c \int_0^1 qe^{-qs}(1-s)^{\frac{1}{\varepsilon}} ds \right\}.$$

Under Assumption 2(ii), $\eta_{dT} \sim \Gamma(\lambda_{d0}, 1)$. We have $\mathbb{E}\left[\eta_{dT}^{1/\varepsilon}\right] = \Gamma(\lambda_{d0} + 1/\varepsilon)/\Gamma(\lambda_{d0})$.

D.1.2 Proofs

Uniform pricing Given ξ_{dT}, the revenue under uniform prices p_d is

$$R_u^c(p_d, \xi_{dT}) = \mathbb{E}[p_d D_{dT}(0, \tau_{C_{dT}} \wedge 1; p_d)|\xi_{dT}],$$

where $\tau_C = \inf\{t : D_{dT}(0,t;p_d) \geq C_{dT}\}$ is the stopping time of selling out all C_{dT} seats. Then,

$$R_u^c(\xi_{dT}) = \max_{p>0} R_u^c(p, \xi_{dT})$$
$$= \max_{p>0} p\mathbb{E}[D(\xi_{dT} p^{-\varepsilon}) \wedge C_{dT}|\xi_{dT}].$$

We obtain the result by defining $q = \xi_{dT} p^{-\varepsilon}$ and integrate over ξ_{dT}.
Without pre-allocation of capacities among intermediate and final destinations, given (p_a, p_b), (ξ_{aT}, ξ_{bT}), and using the first statement of Theorem 4.2, we have:

$$\begin{aligned} R_u^c(p_a, p_b, \xi_{aT}, \xi_{bT}) &= \mathbb{E}\left[\mathbb{E}\left[p_a D(\xi_{aT} p_a^{-\varepsilon}) + p_b D(\xi_{bT} p_b^{-\varepsilon})\Big| \left(D(\xi_{aT} p_a^{-\varepsilon} \xi_{bT} p_b^{-\varepsilon}) \wedge C_T\right)\right]\right] \\ &= \frac{\xi_{aT} p_a^{1-\varepsilon} + \xi_{bT} p_b^{1-\varepsilon}}{\xi_{aT} p_a^{-\varepsilon} + \xi_{bT} p_b^{-\varepsilon}} \mathbb{E}\left[D(\xi_{aT} p_a^{-\varepsilon} + \xi_{bT} p_b^{-\varepsilon}) \wedge C_T\right]. \end{aligned}$$

Then, the optimal revenue under uniform pricing without pre-allocation is achieved when $p_a = p_b$ and therefore:

$$R_u^c(\xi_{aT}, \xi_{bT}) = \max_{p>0} p \mathbb{E}\left[D((\xi_{aT} + \xi_{bT}) p^{-\varepsilon}) \wedge C_T\right]$$

Full dynamic pricing. Denote by $V_k(t, p_d)$ the expected revenue when there remains k vacant seats before the departure and the current seat is priced at p_d at time $1-t$. From $1-t$ to $1-t+\Delta t$, the probability of selling one seat is $b_T(1-t)\xi_{dT} p_d^{-\varepsilon} \Delta t + o(\Delta t)$ and generates p_d revenue if one seat is sold. With probability $o(\Delta t)$, more than one seats are sold. Then, following Gallego and Van Ryzin (1994) (Section 2.2.1 on page 1004), we have:

$$\begin{aligned} V_k^*(t) = \max_{p_d>0} \Big\{ & b_T(1-t)\xi_{dT} p_d^{-\varepsilon} \Delta t \left(p_d + V_{k-1}^*(t-\Delta t)\right) \\ & + [1 - b_T(1-t)\xi_{dT} p_d^{-\varepsilon} \Delta t] V_k^*(t-\Delta t) + o(\Delta t) \Big\}. \end{aligned} \tag{23}$$

Letting $\Delta t \to 0$, this equation shows that V_k^* is continuous. Further, by considering $(V_k^*(t) - V_k^*(t-\Delta t))/\Delta t$ and letting $\Delta t \to 0$, we obtain that V_k^* is differentiable, with[18]

$$V_k^{*\prime}(t) = \max_{p_d>0} b_T(1-t)\xi_{dT} p_d^{-\varepsilon}\left[p_d + V_{k-1}^*(t) - V_k^*(t)\right] \tag{24}$$

with boundary conditions $V_k^*(0) = 0$ for any $k = 1, ..., C_{dT}$ and $V^*(t, 0) = 0$ for any $t \in [0,1]$. As a consequence, the optimal price p_{tk}^* can be obtained from the first-order condition of the right-hand side of (24):

$$p_{tk}^* = \frac{\varepsilon}{\varepsilon - 1}\left[V_k^*(t) - V_{k-1}^*(t)\right]. \tag{25}$$

[18] For conditions that enable to interchange $\lim_{\Delta t \to 0}$ and max, we refer to Brémaud (1981) for details.

By plugging p^*_{tk} into (24) and using $B_T(t,1) = \int_t^1 b_T(s)ds$ (where we let $B_T(t,t') := \int_t^1 b_T(s)ds$), we obtain:

$$V_k^{*\prime}(t) = \partial_1 B_T(1-t,1)\frac{\xi_{dT}}{\varepsilon-1}\left(1-\frac{1}{\varepsilon}\right)^{\varepsilon}\left[V_k^*(t) - V_{k-1}^*(t)\right]^{1-\varepsilon}, \tag{26}$$

where $\partial_j B_T$ denotes the derivative of B_T with respect to its j-th argument. We now prove by induction on k that

$$V_k^*(t) = \alpha_{k,f}^c[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}} \tag{27}$$

for all $k \in \{0,...,C_{dT}\}$, with $\alpha_f^c(0) = 0$ and $\alpha_{k,f}^c = (\alpha_{k,f}^c - \alpha_{k-1,f}^c)^{1-\varepsilon}\left(1-\frac{1}{\varepsilon}\right)^{\varepsilon-1}$.

The result holds for $k=0$ since $V_0^*(t) = 0$. Next, suppose that (27) holds for $k-1 \geq 0$ and let us show that the result holds for k. By plugging this solution for $k-1$ into the differential equation (24), we obtain:

$$V_k^{*\prime}(t) = \partial_1 B_T(1-t,1)\frac{\xi_{dT}}{\varepsilon-1}\left(1-\frac{1}{\varepsilon}\right)^{\varepsilon}\left[V_k^*(t) - \alpha_{k-1,f}^c[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}}\right]^{1-\varepsilon}, \tag{28}$$

with $V_k^*(0) = 0$. We can check that $V_k^*(t) = \alpha_{k,f}^c[\xi_{dT}B_T(1-t,1)]^{1/\varepsilon}$ is a solution to (28). To show uniqueness, let $\phi(v,z) = \frac{1}{\varepsilon-1}\left(1-\frac{1}{\varepsilon}\right)^{\varepsilon}\left[v - \alpha_{k-1,f}^c z^{1/\varepsilon}\right]^{1-\varepsilon}$. Consider the diffeomorphism $z(t) = \xi_{dT}B_T(1-t,1)$ and define $\bar{V}_k^*(z) = V_k^*(t(z))$. Then, (28) can be written as

$$\bar{V}_k^{*\prime}(z) = \phi(\bar{V}_k^*(z), z), \tag{29}$$

with $\bar{V}_k^*(0) = 0$. It is enough to prove that $\bar{V}_k^*$ is the unique solution of (29) and we prove this by contradiction. Suppose that there is another differentiable solution $\tilde{V}_k(.)$ different from $\bar{V}_k^*(z) = \alpha_{k,f}^c z^{1/\varepsilon}$. Without loss of generality, $\tilde{V}_k(z_0) > \bar{V}_k^*(z_0)$ for some $z_0 > 0$. Because $\tilde{V}_k(0) = \bar{V}_k^*(0) = 0$, then $z_m = \sup\{z \leq z_0 : \tilde{V}_k(z_0) \leq \bar{V}_k^*(z_0)\}$ exists and $z_m < z_0$. Moreover, $\tilde{V}_k(z_m) = \bar{V}_k^*(z_m)$. Then, (29) implies the contradiction

$$0 < \tilde{V}_k(z_0) - \bar{V}_k^*(z_0) = \int_{z_m}^{z_0}[\phi(\tilde{V}_k(z), z) - \phi(\bar{V}_k^*(z), z)]dz \leq 0,$$

where the second inequality follows from the fact that ϕ is a decreasing function of z and $\tilde{V}_k(s) > \bar{V}_k^*(s)$ for all $s \in (z_m, z_0]$. Finally, we conclude that $\bar{V}_k^*(.)$ is the unique solution. Hence, the result holds for k, and (27) holds. By taking $t=1, k=C_{dT}$ and integrating over ξ_{dT}, we obtain the formula in Section D.

Stopping-time pricing. Denote by $V_k(t, p_d)$ the expected optimal revenue at time $1-t$ when pricing the next seat at p_d and with k remaining seats. In this scenario, prices do not change until the next seat is sold. Define $\tau_{1-t,p_d} = \inf\{s > 0 : D_T(1-t, 1-t+s; p_d) \geq 1\}$. Then,

$$\begin{aligned}\Pr[\tau_{1-t;p_d} > s] &= \Pr[D(1-t, 1-t+s; p_d) = 0] \\ &= \exp\{-B_T(1-t, 1-t+s)\xi_d p_d^{-\varepsilon}\},\end{aligned}$$

and the density of τ_{1-t,p_d} is

$$f_{\tau_{1-t,p_d}}(s) = \xi_d p_d^{-\varepsilon} \partial_2 B_T(1-t, 1-t+s) e^{-B_T(1-t,1-t+s)\xi_d p_d^{-\varepsilon}}. \tag{30}$$

Then, the Bellman equation is

$$\begin{aligned}V_k(t, p_d) &= \mathbb{E}\left[\mathbf{1}_{\tau_{1-t,p_d}<t}\left(p_d + V_{k-1}^*(t - \tau_{1-t,\boldsymbol{p}})\right)\right] \\ &= \int_0^t f_{\tau_{1-t,p_d}}(s)\left(p_d + V_{k-1}^*(t-s)\right) ds \\ &= \int_0^t \xi_d p_d^{-\varepsilon} \partial_2 B_T(1-t, 1-t+s) e^{-B_T(1-t,1-t+s)\xi_d p_d^{-\varepsilon}} \times \left(p_d + V_{k-1}^*(t-s)\right) ds.\end{aligned} \tag{31}$$

Let $V_k^*(t) = \max_{p>0} V_k(t, p)$. We now show by induction that

$$V_k^*(t) = \alpha_{k,s}^c [\xi_{dT} B_T(1-t, 1)]^{\frac{1}{\varepsilon}}, \tag{32}$$

where $\alpha_{0,s}^c = 0$ and

$$\alpha_{k,s}^c = \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 q e^{-sq}(1-s)^{\frac{1}{\varepsilon}} ds \right\}.$$

The result holds for $k = 0$ since $V_0^*(1-t) = 0$. Now, suppose that (32) is true for $k-1 \geq 0$. By using the change of variable $z = B_T(1-t, 1-t+s)/B_T(1-t, 1)$ and applying (32) for $V_{k-1}^*(t)$ in Equation (31), we get

$$\begin{aligned}V_k(t, p) &= \int_0^1 \xi_{dT} B_T(1-t, 1) p^{-\varepsilon} e^{-B_T(1-t,1)\xi_{dT} p^{-\varepsilon} z} \left(p + [\xi_{dT} B_T(1-t, 1)(1-z)]^{\frac{1}{\varepsilon}} \alpha_{k-1,s}^c\right) dz \\ &= [\xi_{dT} B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \left(q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 q e^{-qz}(1-z)^{\frac{1}{\varepsilon}} dz\right),\end{aligned}$$

where $q = \xi_{dT} B_T(1-t, 1) p^{-\varepsilon}$. As a consequence,

$$\begin{aligned}V_k^*(t) &= \max_{p>0} V_k(t, p) \\ &= [\xi_{dT} B_T(1-t, 1)]^{\frac{1}{\varepsilon}} \max_{q>0} \left\{ q^{-\frac{1}{\varepsilon}}(1 - e^{-q}) + \alpha_{k-1,s}^c \int_0^1 q e^{-qz}(1-z)^{\frac{1}{\varepsilon}} dz \right\} \\ &= \alpha_{k,s}^c [\xi_{dT} B_T(1-t, 1)]^{\frac{1}{\varepsilon}},\end{aligned}$$

and (32) is true for k. Thus, (32) holds for all $k \in \{0, ..., C_{dT}\}$. Finally, by taking $t = 1, k = C_{dT}$ and the expectation with respect to $\xi_{dT} | (X_{dT}, W_T)$, we obtain the expression in Appendix D.

Stopping-time pricing with M fares. Denote by $V_k(0; t, p, m)$ (resp. $V_k(1; t, p, m)$) the expected revenue of the firm at time $1 - t$, with a current price p, a remaining capacity k and a remaining number of fares m, if it decides to keep the same price p (resp. to choose a new price). Then, we have the following Bellman equations:

$$\begin{cases} V_k(1; t, p, m) = \max_{p' > 0} \int_0^t f_{\tau_{1-t,p'}}(s) \left[p' + V^*_{k-1}(t - s, p', m - 1) \right] ds, \\ V_k(0; t, p, m) = \int_0^t f_{\tau_{1-t,p}}(s) \left[p + V^*_{k-1}(t - s, p, m) \right] ds, \\ V^*_k(t, p, m) = \max_{d \in \{0,1\}} V_k(d; t, p, m), \end{cases} \tag{33}$$

with initial conditions $V^*_0(t, p, m) = 0$. We show by induction on k that for all $(k, m) \in \{0, ..., C_{dT}\} \times \mathbb{N}$,

$$V^*_k(t, p, m) = \alpha_{k,m}(q(t, p)) \left[\xi_{dT} B_T(1 - t, 1) \right]^{\frac{1}{\varepsilon}}, \tag{34}$$

where $q(t, p) = p^{-\varepsilon} \xi_{dT} B_T(1 - t, 1)$, $\alpha_{k,0}(q) = q^{-\frac{1}{\varepsilon}} \mathbb{E}[D(q) \wedge k]$ and for $m \geq 1$,

$$\alpha_{k,m}(q) = \max \left\{ q \int_0^1 e^{-qz} \left[q^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m \wedge (k-1)}(q(1 - z))(1 - z)^{\frac{1}{\varepsilon}} \right] dz, \right.$$
$$\left. \max_{q' > 0} q' \int_0^1 e^{-q'z} \left[q'^{-\frac{1}{\varepsilon}} + \alpha_{k-1, m-1}(q'(1 - z))(1 - z)^{\frac{1}{\varepsilon}} \right] dz. \right\}.$$

Because for any $m \geq k$ and $d \in \{0, 1\}$, we have $V_k(d; t, p, m) = V_k(d; t, p, k)$, it suffices to prove the result for $m \leq k$. The result holds for $k = m = 0$ since $V^*_0(t, p, m) = 0$. Now, suppose that (34) holds for $k - 1 \geq 0$ and all $m \leq k - 1$. If $m = 0$, the price cannot be changed anymore, so $V^*_k(t, p, m)$ is simply the revenue with price p from $1 - t$ to 1, and (34) holds.

If $m \geq 1$, we have, by Equations (30), (33), the change of variable $z = B_T(1 - t, 1 -$

$t+s)/B_T(1-t,1)$ and the induction hypothesis,

$$
\begin{aligned}
&V_k(0;t,p,m)\\
=&\int_0^t f_{\tau_{1-t,p}}(s)\left[p+V^*_{k-1}(t-s,p,m\wedge(k-1))\right]ds\\
=&\int_0^t \xi_{dT}p^{-\varepsilon}\partial_2 B_T(1-t,1-t+s)e^{-\xi_{dT}p^{-\varepsilon}B_T(1-t,1-t+s)}\\
&\left[p+\alpha_{k-1,m\wedge(k-1)}(\xi_{dT}B_T(1-t+s,1)p^{-\varepsilon})[\xi_{dT}B_T(1-t+s,1)]^{\frac{1}{\varepsilon}}\right]ds\\
=&\int_0^1 \xi_{dT}p^{-\varepsilon}B_T(1-t,1)e^{-\xi_{dT}p^{-\varepsilon}B_T(1-t,1)z}\\
&\left[p+\alpha_{k-1,m\wedge(k-1)}(\xi_{dT}B_T(1-t,1)p^{-\varepsilon}(1-z))[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}}(1-z)^{\frac{1}{\varepsilon}}\right]dz\\
=&[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}}\int_0^1 q(t,p)e^{-q(t,p)z}\left[q(t,p)^{-\frac{1}{\varepsilon}}+\alpha_{k-1,m\wedge(k-1)}(q(t,p)(1-z))(1-z)^{\frac{1}{\varepsilon}}\right]dz,
\end{aligned}
\tag{35}
$$

With the same reasoning, we also obtain

$$
\begin{aligned}
&V_k(1;t,p,m)\\
=&\max_{p'>0}\int_0^t f_{\tau_{1-t,p'}}(s)\left[p'+V^*_{k-1}(t-s,p',m-1)\right]ds\\
=&[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}}\max_{q>0}\int_0^1 qe^{-qz}\left[q^{-\frac{1}{\varepsilon}}+\alpha_{k-1,m-1}(q(1-z))(1-z)^{\frac{1}{\varepsilon}}\right]dz.
\end{aligned}
$$

Then,

$$
\begin{aligned}
V^*_k(t,p,m)&=\max_{d\in\{0,1\}}V_k(d;t,p,m)\\
&=\alpha_{k,m}(q(t,p))[\xi_{dT}B_T(1-t,1)]^{\frac{1}{\varepsilon}},
\end{aligned}
$$

Thus, (34) holds for k, and hence for all $k\in\{0,...,C_{dT}\}$. By setting $t=0$ and optimizing $V^*_k(t,p,m)$ over p (or equivalently over $q(t,p)$) and taking the expectation with respect to $\xi_{dT}|(X_{dT},W_T)$, we obtain the desired expression in Appendix D.

Stopping-time pricing with M increasing prices. The reasoning is very similar to the previous case. The only change in (33) is in the formula of $V_k(1;t,p,m)$: the maximization is now over $p'\geq p$ rather than $p'\geq 0$, since the new price has to be higher than the current one. Then, following a similar strategy by induction, we get

$$
V^*_k(t,p,m)=\alpha^+_{k,m}(q(t,p))\left[\xi_{dT}B_T(1-t,1)\right]^{\frac{1}{\varepsilon}},
$$

where $\alpha^+_{k,0}(q) = \alpha_{k,0}(q)$ and

$$\alpha^+_{k,m}(q) = \max\left\{q\int_0^1 e^{-qz}\left[q^{-\frac{1}{\varepsilon}} + \alpha^+_{k-1,m\wedge(k-1)}(q(1-z))(1-z)^{\frac{1}{\varepsilon}}\right]dz,\right.$$
$$\left.\max_{q'\in(0,q]} q'\int_0^1 e^{-q'z}\left[q'^{-\frac{1}{\varepsilon}} + \alpha^+_{k-1,m-1}(q'(1-z))(1-z)^{\frac{1}{\varepsilon}}\right]dz.\right\}.$$

We obtain the result by taking $t = 0$, $k = C_{dT}$ and defining $\alpha^c_{C_{dT},sM+} = \max_{q>0}\alpha^+_{C_{dT},M}(q)$.

Intermediate-K stopping-time pricing The proof is the same as that for (32) except for the initial value because the firm must apply uniform pricing whenever there remain $C_{dT}(1-K\%)$ seats. Thus, the Bellman equation and the updating of the constants $\alpha^c_{iK,k}$ take the same form as under the stopping-time pricing strategy in (32) for $k \geq C_{dT}(1-K\%)$. The initial value becomes $\alpha^c_{iK,C_{dT}(1-K\%)}$, which comes from the optimal uniform pricing with $C_{dT}(1-K\%)$ seats.

D.2 Incomplete Information

Hereafter, we denote by $g_{\lambda,\mu}$ the density of the $\Gamma(\lambda,\mu)$ distribution.

D.2.1 Formulas

Uniform pricing $\alpha^i_u = \max_{q>0}\left\{\int_{\mathbb{R}^+} q^{-\frac{1}{\varepsilon}}\mathbb{E}[D(qz)\wedge C_{dT}]f(z)dz\right\}$. Under Assumption 2(ii), $\alpha^i_u = \max_{q>0}\left\{\int_{\mathbb{R}^+} q^{-\frac{1}{\varepsilon}}\mathbb{E}[D(qz)\wedge C_{dT}]g_{\lambda_0,1}(z)dz\right\}$.

Full-dynamic pricing Under Assumption 2, $\alpha^i_f = \alpha^i_{C_{dT},f}(\lambda_{d0})$, where $\alpha^i_{0,f}(\lambda) = 0$ for any $\lambda > 0$ and for all $k \in \{1, ..., C_{dT}\}$,

$$\alpha^i_{k,f}(\lambda) = \lambda\left(1-\frac{1}{\varepsilon}\right)^{\varepsilon-1}\left[-\alpha^i_{k-1,f}(\lambda+1) + (1+\frac{1}{\lambda\varepsilon})\alpha^i_{k,f}(\lambda)\right]^{1-\varepsilon}.$$

Stopping-time pricing $\alpha^i_s = \alpha^i_{C_{dT},s}(f)$, where $\alpha^i_{0,s}(f) = 0$ and for any $k \in \{1, .., C_{dT}\}$,

$$\alpha^i_{k,s}(f) = \max_{q>0} q\int_0^1\left[q^{-1/\varepsilon} + (1-u)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,s}(T(f;qu))\right]\left[\int_0^\infty ze^{-quz}f(z)dz\right]du$$

and $T(f;q)$ is a transformation of density function f defined in Lemma D.1 below. Under Assumption 2(ii), $\alpha^i_s = \alpha^i_{C_{dT},s}(\lambda_{d0})$, where $\alpha^i_{0,s}(\lambda) = 0$ for $\lambda > 0$, and for all

$k \in \{1, ..., C_{dT}\}$,

$$\alpha^i_{k,s}(\lambda) = \max_{q>0} q \int_0^1 \frac{\lambda}{(1+qs)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + \left(\frac{1-s}{1+qs}\right)^{\frac{1}{\varepsilon}} \alpha^i_{k-1,s}(\lambda+1) \right] ds.$$

Stopping-time pricing with M fares $\alpha^i_{M,s}(M, f) = \alpha^i_{sM}(f)$ where $\alpha^i_{sM}(f) = \max_{q>0} c_{C_{dT},M}(q, f)$ and for all k, $c_{k,0}(q, f) = q^{-\frac{1}{\varepsilon}} \int \mathbb{E}[D(qz) \wedge k] f(z) dz$ and

$$c_{k,m}(q, f) = \max \Big\{ q \int_0^1 \int z e^{-qzu} f(z) dz \Big[q^{-1/\varepsilon} + c_{k-1,m\wedge(k-1)}(q(1-u), T(f; qu)) (1-u)^{\frac{1}{\varepsilon}} \Big] du, \ \max_{q'>0} q' \int_0^1 \int z e^{-q'zu} f(z) dz \Big[q'^{-1/\varepsilon} + c_{k-1,m-1}(q'(1-u), T(f; q'u))(1-u)^{\frac{1}{\varepsilon}} \Big] du \Big\}$$

for any $m \in \{1, ..., k\}$, T being the same transform as in the case of stopping-time pricing. Further under Assumption 2(ii), $\alpha^i_{sM}(M, \lambda_{d0}) = \alpha^i_{sM}(\lambda_0)$, where $\alpha^i_{sM}(\lambda) = \max_{q>0} c_{C_{dT},M}(q, \lambda)$ with, for all k, $c_{k,0}(q, \lambda) = q^{-\frac{1}{\varepsilon}} \int \mathbb{E}[D(qz) \wedge k] g_{\lambda,1}(z) dz$ and for all $k \in \{1, ..., C_{dT}\}$ and all $m \in \{1, ..., k\}$,

$$c_{k,m}(q, \lambda) = \max \Bigg\{ q \int_0^1 \frac{\lambda}{(1+qu)^{\lambda+1}} \left[q^{-\frac{1}{\varepsilon}} + c_{k-1,m\wedge(k-1)} \left(\frac{q(1-u)}{1+qu}, \lambda+1 \right) \left(\frac{1-u}{1+qu} \right)^{\frac{1}{\varepsilon}} \right] du, \max_{q'>0} q' \int_0^1 \frac{\lambda}{(1+q'u)^{\lambda+1}} \left[q'^{-\frac{1}{\varepsilon}} + c_{k-1,m-1} \left(\frac{q'(1-u)}{1+q'u}, \lambda+1 \right) \left(\frac{1-u}{1+q'u} \right)^{\frac{1}{\varepsilon}} \right] du \Bigg\}.$$

Stopping-time pricing with M increasing fares $\alpha^i_{sM+}(M, f) = \alpha^i_{C_{dT},sM+}(f)$, where $\alpha^i_{C_{dT},sM+}(f) = \max_{q>0} c^+_{C_{dT},M}(q, f)$ with, for any $k \in \{0, ..., C_{dT}\}$, $c^+_{k,0}(q, f) = c_{k,0}(q, f)$ and for any $m \geq 1$,

$$c^+_{k,m}(q, f) = \max \Big\{ q \int_0^1 \int z e^{-qzu} f(z) dz \Big[q^{-1/\varepsilon} + c^+_{k-1,m\wedge(k-1)}(q(1-u), T(f; qu)) (1-u)^{\frac{1}{\varepsilon}} \Big] du, \ \max_{q' \in (0,q]} q' \int_0^1 \int z e^{-q'zu} f(z) dz \Big[q'^{-1/\varepsilon} + c^+_{k-1,m-1}(q'(1-u), T(f; q'u))(1-u)^{\frac{1}{\varepsilon}} \Big] du \Big\}$$

Under Assumption 2(ii), we have $R^i_{sM+}(M, \lambda_{d0}) = \alpha^i_{sM+}(\lambda_0)$, where $\alpha^i_{sM+}(\lambda) = \max_{q>0} c^+_{C_{dT},M}(q, \lambda)$ with $c^+_{k,m}(q, \lambda) = c^+_{k,m}(q, g_{\lambda,1})$ as defined above. Further, we have the following simplifications:

$$c^+_{k,m}(q,\lambda) = \max\left\{q\int_0^1 \frac{\lambda}{(1+qu)^{\lambda+1}}\left[q^{-\frac{1}{\varepsilon}} + c^+_{k-1,m\wedge(k-1)}\left(\frac{q(1-u)}{1+qu},\lambda+1\right)\left(\frac{1-u}{1+qu}\right)^{\frac{1}{\varepsilon}}\right]du,\right.$$
$$\left.\max_{q'\in(0,q]} q'\int_0^1 \frac{\lambda}{(1+q'u)^{\lambda+1}}\left[q'^{-\frac{1}{\varepsilon}} + c^+_{k-1,m-1}\left(\frac{q'(1-u)}{1+q'u},\lambda+1\right)\left(\frac{1-u}{1+q'u}\right)^{\frac{1}{\varepsilon}}\right]du\right\}.$$

Intermediate-K stopping-time pricing $\alpha^i_{iK} = \alpha^i_{C_{dT},iK}(f)$, where $\alpha^i_{C_{dT}(1-K\%),iK}(f) = \max_{q>0}\left\{\int_{\mathbb{R}^+} q^{-\frac{1}{\varepsilon}}\mathbb{E}[D(qz)\wedge(C_{dT}(1-K\%))]f(z)dz\right\}$ and for any $k > C_{dT}(1-K\%)$,

$$\alpha^i_{k,iK}(f) = \max_{q>0} q\int_0^1\left[q^{-1/\varepsilon} + (1-u)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,iK}(T(f;qu))\right]\left[\int_0^\infty ze^{-quz}f(z)dz\right]du.$$

Under Assumption 2(ii), $\alpha^i_s = \alpha^i_{C_{dT},iK}(\lambda_0)$, where

$$\alpha^i_{C_{dT}(1-K\%),iK}(\lambda) = \max_{q>0}\left\{\int_{\mathbb{R}^+} q^{-\frac{1}{\varepsilon}}\mathbb{E}[D(qz)\wedge(C_{dT}(1-K\%))]g_{\lambda,1}(z)dz\right\}$$

for $\lambda > 0$, and for all $k > C_{dT}(1-K\%)$,

$$\alpha^i_{k,iK}(\lambda) = \max_{q>0}\left\{q\int_0^1 \frac{\lambda}{(1+qs)^{\lambda+1}}\left[q^{-\frac{1}{\varepsilon}} + \left(\frac{1-s}{1+qs}\right)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,iK}(\lambda+1)\right]ds.\right\}.$$

D.2.2 Proofs

Denote $Y_{dT} = \exp\{X'_{dT}\beta_0\}g_0(W_T)$. Denote the density function of ξ_{dT} by f. Under Assumption 2(ii), f is a gamma density $\Gamma(\lambda_{d0}, Y^{-1}_{dT})$.

Uniform pricing. We have:

$$\begin{aligned} R^i_u &= \max_{p>0} R^i_u(p;\varepsilon,f) \\ &= \max_{p>0} p\int_{z>0}\mathbb{E}[D(p^{-\varepsilon}z)\wedge C_{dT}]f(z)dz. \end{aligned}$$

By the change of variable $q = Y_{dT}p^{-\varepsilon}$, we obtain the desired formula.
Now, given (p_a,p_b) and (C_{aT},C_{bT}), the total revenue generated by train T is:

$$\begin{aligned} &\mathbb{E}\left[R_T(p_a,p_b;C_{aT},C_{bT})|W_T\right] \\ =&\sum_{d=a,b} p_d\int_{z>0}\mathbb{E}\left[D(p_d^{-\varepsilon}\exp\{X'_{dT}\beta_0\}g_0(W_T)z)\wedge C_{dT}\right]g_{\lambda_{d0},1}(z)dz, \end{aligned} \tag{36}$$

where $g_{\lambda_{d0},1}(z)$ is the density of a $\Gamma(\lambda_{d0},1)$. Then, we obtain (6) by maximizing the revenue in (36) over all possible allocations (C_{aT}, C_{bT}) subject to $C_{aT}+C_{bT}=C_T$. Without pre-allocation of capacities among intermediate and final destinations, we obtain:

$$\begin{aligned}\mathbb{E}\left[R_T(p_a,p_b;C_T)|W_T\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{d=a,b} p_d D_{dT}\Big|(D_{aT}+D_{bT})\wedge C_T, z_a, z_b\right]\right] \\ &= \mathbb{E}\left[\left(D\left(\sum_{d=a,b} p_d^{-\varepsilon}\exp\{X'_{dT}\beta_0\}g_0(W_T)z_d\right)\wedge C_T\right)\frac{\sum_{d=a,b} p_d^{1-\varepsilon}\exp\{X'_{dT}\beta_0\}g_0(W_T)z_d}{\sum_{d=a,b} p_d^{-\varepsilon}\exp\{X'_{dT}\beta_0\}g_0(W_T)z_d}\right]\end{aligned}$$

where z_a and z_b follows $\Gamma(\lambda_{a0},1)$ and $\Gamma(\lambda_{b0},1)$, respectively, and are independent. Then, the optimal revenue is obtained by maximizing $\mathbb{E}\left[R_T(p_a,p_b;C_T)|W_T\right]$ over $(p_a,p_b)\in\mathbb{R}_+^2$.

Full dynamic pricing. Define $V_k(t,p,f)$ as the expected revenue at time $1-t$ when there remains k vacant seats before the departure, the current seat is priced at p and the density of ξ_{dT}, given the current information, is f. Let also $V_k^*(t,f)=\max_{p>0}V_k(t,p,f)$. When $\eta_T\sim\Gamma(\lambda,\mu)$, we use respectively $V_k(t,p,\lambda,\mu)$ and $V_k^*(t,\lambda,\mu)$ instead of $V_k(t,p,g_{\lambda,\mu})$ and $V_k^*(t,g_{\lambda,\mu})$.

Between $1-t$ and $1-t+\Delta t$, if one seat is sold, which occurs with probability $\xi_{dT}p^{-\varepsilon}\partial_1 B_T(1-t,1)\Delta t+o(\Delta t)$, the posterior cdf of ξ_{dT}, $F_1(.;\Delta t)$ satisfies

$$F_1(\xi;\Delta t)\propto[p^{-\varepsilon}\partial_1 B_T(1-t,1)\xi\Delta t+o(\Delta t)]\xi^{\lambda-1}e^{-\mu\xi},$$

and the corresponding density is

$$f_1(\xi;\Delta t)=\xi^{\lambda}e^{-\mu\xi}\frac{\mu^{\lambda+1}}{\Gamma(\lambda+1)}+o(\Delta t).$$

As $\Delta t\to 0$, the posterior density converges to $g_{\lambda+1,\mu}$. If the seat is not sold between $1-t$ and $1-t+\Delta t$, then the posterior cdf of η_T is

$$F_0(\xi;\Delta t)\propto\xi^{\lambda-1}\exp(-\mu(t,\Delta t,p)\xi),$$

where $\mu(t,\Delta t,p)=\mu+p^{-\varepsilon}B_T(1-t,1-t+\Delta t)$. Therefore, the posterior density is $g_{\lambda,\mu(t,\Delta t,p)}$. Then, the Bellman equation can be written as:

$$\begin{aligned}V_k(t,p,\lambda,\mu)=\int\Big\{&[p^{-\varepsilon}\xi\partial_1 B_T(1-t,1)\Delta t+o(\Delta t)]\times\left[p+V_{k-1}^*(t-\Delta t,f_1(.;\Delta t))\right] \\ &+\left[1-p^{-\varepsilon}\xi\partial_1 B_T(1-t,1)\Delta t-o(\Delta t)\right]\times V_k^*(t-\Delta t,\lambda,\mu(p,t,\Delta))\Big\}g_{\lambda,\mu}(\xi)d\xi.\end{aligned}$$

Then,

$$V_k(t,p,\lambda,\mu) = V_k^*(t-\Delta t,\lambda,\mu) + \int \Big\{[p^{-\varepsilon}\xi\partial_1 B_T(1-t,1)\Delta t + o(\Delta t)] \times [p + V_{k-1}^*(t-\Delta t, f_1(.;\Delta t))] + [V_k^*(t-\Delta t,\lambda,\mu(p,t,\Delta t)) - V_k^*(t-\Delta t,\lambda,\mu)] - V_k^*(t-\Delta t,\lambda,\mu(t,\Delta,p))[p^{-\varepsilon}\xi\partial_1 B_T(1-t,1)\Delta t + o(\Delta t)]\Big\} g_{\lambda,\mu}(\xi)d\xi.$$

Then, using $V_k^*(t,\lambda,\mu) = \max_{p>0} V_k(t,p,\lambda,\mu)$ and letting $\Delta t \to 0$, we obtain:[19]

$$\begin{aligned}
&\partial_1 V_k^*(t,\lambda,\mu)\\
&= \max_{p>0} \int \Big\{ p^{-\varepsilon}\xi\partial_1 B_T(1-t,1)\left[p + V_{k-1}^*(t,\lambda+1,\mu) - V_k^*(t,\lambda,\mu)\right]\\
&\qquad + \lim_{\Delta t\to 0} \frac{V_k^*(t-\Delta t,\lambda,\mu(t,\Delta t,p)) - V_k^*(t-\Delta t,\lambda,\mu)}{\Delta t}\Big\} g_{\lambda,\mu}(\xi)d\xi\\
&= \partial_1 B_T(1-t,1)\max_{p>0}\int\Big\{p^{-\varepsilon}\xi\left[p + V_{k-1}^*(t,\lambda+1,\mu) - V_k^*(t,\lambda,\mu)\right] + \partial_3 V_k^*(t,\lambda,\mu)p^{-\varepsilon}\Big\} g_{\lambda,\mu}(\xi)d\xi\\
&= \partial_1 B_T(1-t,1)\max_{p>0}\Big\{p^{-\varepsilon}\frac{\lambda}{\mu}\left[p + V_{k-1}^*(t,\lambda+1,\mu) - V_k^*(t,\lambda,\mu)\right] + \partial_3 V_k^*(t,\lambda,\mu)p^{-\varepsilon}\Big\}.
\end{aligned}$$

Solving for the optimal price, we then obtain:

$$\partial_1 V_k^*(t,\lambda,\mu) = \left[\frac{\varepsilon}{\varepsilon-1}\right]^{-\varepsilon}\frac{\lambda}{\mu(\varepsilon-1)}\partial_1 B_T(1-t,1) \times \left[-V_{k-1}^*(t,\lambda+1,\mu) + V_k^*(t,\lambda,\mu) - \frac{\mu}{\lambda}\partial_3 V_k^*(t,\lambda,\mu)\right]^{1-\varepsilon}.$$

Letting $z(t) = B_T(1-t,1)$ and $\bar{V}^*(z(t),\lambda,\mu) = V^*(t,\lambda,\mu)$, we obtain:

$$\partial_1 \bar{V}_k^*(z,\lambda,\mu) = \left[\frac{\varepsilon}{\varepsilon-1}\right]^{-\varepsilon}\frac{\lambda}{\mu(\varepsilon-1)}\left[-\bar{V}_{k-1}^*(z,\lambda+1,\mu) + \bar{V}_k^*(z,\lambda,\mu) - \frac{\mu}{\lambda}\partial_3\bar{V}_k^*(z,\lambda,\mu)\right]^{1-\varepsilon}. \tag{37}$$

We prove by induction on k that for all $k \in \{0,...,C_{dT}\}$.

$$\bar{V}_k^*(z,\lambda,\mu) = \left(\frac{z}{\mu}\right)^{\frac{1}{\varepsilon}}\alpha_{k,f}^i(\lambda), \tag{38}$$

where $\alpha_f^i(0,\lambda) = 0$ and for $k \geq 1$,

$$\alpha_{k,f}^i(\lambda) = \lambda\left(1 - \frac{1}{\varepsilon}\right)^{\varepsilon-1}\left[-\alpha_{k-1,f}^i(\lambda+1) + \left(1 + \frac{1}{\lambda\varepsilon}\right)\alpha_{k,f}^i(\lambda)\right]^{1-\varepsilon}.$$

[19] For conditions that enable to interchange $\lim_{\Delta t\to 0}$ and max, we refer to Brémaud (1981) for details.

The result holds for $k = 0$ since $V_0^*(z, \lambda, \mu) = 0$. Suppose that (38) holds for $k-1$. Then, (37) and the induction hypothesis yield

$$\partial_1 \bar{V}_k^*(z, \lambda, \mu) = \left[\frac{\varepsilon}{\varepsilon - 1}\right]^{1-\varepsilon} \frac{\lambda}{\mu(\varepsilon - 1)} \Bigg[- \left(\frac{z}{\mu}\right)^{\frac{1}{\varepsilon}} \alpha^i_{k-1,f}(\lambda + 1) + \bar{V}_k^*(z, \lambda, \mu) - \frac{\mu}{\lambda} \partial_3 \bar{V}_k^*(z, \lambda, \mu) \Bigg]^{1-\varepsilon}. \tag{39}$$

The function $(z, \lambda, \mu) \mapsto \alpha^i_{k,f}(\lambda)\,(z/\mu)^{1/\varepsilon}$ is a solution to (39). We now show that $\bar{V}_k^*(z, \lambda, \mu)$ is equal to this solution. First, note that $V_k^*(t, \lambda, \mu)$ remains unchanged if the distribution of $B_T(t, t')\xi$ remains unchanged. Now,

$$B_T(t, t')\xi = (B_T(t, t')/\delta) \times (\delta\xi),$$

with $\delta\xi \sim \Gamma(\lambda, \mu/\delta)$. Hence, $V_k^*(t, \lambda, \mu)$ remains unchanged if we replace μ by μ/δ and $z(t)$ by $z(t)/\delta$. Given the definition of $\bar{V}_k^*(z, \lambda, \mu)$, this implies $\bar{V}_k^*(z/\delta, \lambda, \mu/\delta) = \bar{V}_k^*(z, \lambda, \mu)$ for all $\delta > 0$. Then, to prove the induction step, we only need to show that $V(x) := V_k^*(x, \lambda, 1)$ satisfies $V(x) = \alpha^i_{k,f}(\lambda) x^{1/\varepsilon}$. By Equation (39),

$$V'(x) = \left[\frac{\varepsilon}{\varepsilon - 1}\right]^{1-\varepsilon} \frac{\lambda}{\varepsilon - 1} \left[-x^{\frac{1}{\varepsilon}} \alpha^i_{k-1,f}(\lambda + 1) + V(x) + \frac{x}{\lambda} V'(x) \right]^{1-\varepsilon}, \tag{40}$$

with initial condition $V(0) = 0$. Suppose that (40) has two distinct solutions V_1, V_2 and let x_0 be such that $V_1(x_0) \neq V_2(x_0)$, say $V_1(x_0) > V_2(x_0)$. Define $x_m = \sup\{x \leq x_0 : V_1(x) \leq V_2(x)\}$. Because $V_1(0) = V_2(0)$ and $V_1(x_0) > V_2(x_0)$, we have $0 \leq x_m < x_0$ and $V_1(x) > V_2(x)$ for $x \in (x_m, x_0]$. Moreover, because both solutions are continuous, $V_1(x_m) = V_2(x_m)$. According to (40), because $\varepsilon > 1$, as long as $V_1(x) > V_2(x)$, we have $V_1'(x) < V_2'(x)$. Then,

$$V_1(x_0) - V_2(x_0) = \int_{x_m}^{x_0} [V_1'(x) - V_2'(x)]\, dx < 0,$$

which contradicts $V_1(x_0) > V_2(x_0)$. Hence, $V(x) = \alpha^i_{k,f}(\lambda) x^{1/\varepsilon}$, and the induction step holds. Thus, (38) is satisfied for $k \in \{0, ..., C_{dT}\}$. Finally, we obtain the result in Appendix D by taking $t = 0$ and $k = C_{dT}$.

Stopping-time pricing The difference from the stopping-time pricing under complete information is that the firm updates in a Bayesian way its belief on the distribution of ξ_{dT}. Even if the firm continuously updates its belief, only moments where a sale occurs matter, since this is the time where it can decide to change its prices. Thus, starting at time $1-t$, we can focus on time $1-t+\tau_{t,p}$. The next lemma characterizes the corresponding posterior distribution of ξ_{dT}.

Lemma D.1 *Suppose that the density function of ξ_{dT} at time $1-t$ is f and the firm prices the next seat at p. Then, the posterior distribution of $\xi_{dT}|\tau_{1-t;p}=s$ is $T(f;q(t,s,p))$, with $q(t,s,p)=p^{-\varepsilon}B_T(1-t,1-t+s)$ and*

$$T(f;u)(z)=\frac{ze^{-uz}f(z)}{\int ze^{-uz}f(z)dz}.$$

$\xi_d p_d^{-\varepsilon}\partial_2 B_T(1-t,1-t+s)e^{-B_T(1-t,1-t+s)\xi_d p_d^{-\varepsilon}}$.

Proof: As Equation (30) shows, given $\xi_{dT}=z$, the density function of $\tau_{1-t;p}$ is

$$f_{\tau_{1-t,p}|\xi_{dT}}(s|z)=p^{-\varepsilon}z\partial_2 B_T(1-t,1-t+s)e^{-zq(t,s,p)}. \tag{41}$$

Then, the joint distribution of $(\tau_{1-t,p},\xi_{dT})$ is

$$f_{\tau_{1-t,p},\xi_{dT}}(s,z)=p^{-\varepsilon}z\partial_2 B_T(1-t,1-t+s)e^{-q(t,s,p)z}f(z)$$

The result follows.

Now, using the same notation as in the full dynamic pricing case above and the same arguments as in proof of (31), we have

$$V_k(t,p,f)=\int_0^t f_{\tau_{1-t,p}}(s)\Big[p+V^*_{k-1}(t-s;T(f;q(t,s,p)))\Big]ds.$$

and

$$V^*_k(t,f)=\max_{p>0}\int_0^t f_{\tau_{1-t,p}}(s)\left[p+V^*_{k-1}(t-s;T(f;q(t,s,p)))\right]ds. \tag{42}$$

We now prove by induction on k that for all $k\in\{0,...,C_{dT}\}$,

$$V^*_k(t;f)=[B_T(1-t,1)]^{\frac{1}{\varepsilon}}\,\alpha^i_{k,s}(f). \tag{43}$$

where $\alpha^i_s(0,f)=0$ and for all $k\in\{1,..,C_{dT}\}$,

$$\alpha^i_{k,s}(f)=\max_{q>0}q\int_0^1\left[q^{-1/\varepsilon}+(1-u)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,s}(T(f;qu))\right]\int_0^\infty ze^{-quz}f(z)dzdu.$$

The result holds for $k=0$ since $V_0^*(t;f)=0$. Suppose that it holds for $k-1\geq 0$. First, by (41), we have

$$f_{\tau_{1-t,p},B_T}(s,z)=\int_0^\infty p^{-\varepsilon}z\partial_2 B_T(1-t,1-t+s)e^{-q(t,s,p)z}f(z)dz. \tag{44}$$

Using (42), we obtain

$$\begin{aligned}
V_k^*(t,f) =& \max_{p>0}\int_0^t f_{\tau_{1-t,p}}(s)\left\{p+[B_T(1-t+s,1)]^{\frac{1}{\varepsilon}}\times\alpha^i_{k-1,s}(T(f;q(t,s,p)))\right\}ds \\
=& \max_{p>0}\int_0^1\left[\int_0^\infty q(t,p)ze^{-q(t,p)uz}f(z)dz\right] \\
& \times\left\{p+[B_T(1-t,1)(1-u)]^{\frac{1}{\varepsilon}}\,\alpha^i_{k-1,s}(T(f;q(t,p)u))\right\}du \\
=& [B_T(1-t,1)]^{\frac{1}{\varepsilon}}\max_{q>0}q\int_0^1\left[\int_0^\infty ze^{-quz}f(z)dz\right] \\
& \times\left[q^{-1/\varepsilon}+(1-u)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,s}(T(f;qu))\right]du.
\end{aligned}$$

The second equality follows using the change of variable $u=B_T(1-t,1-t+s)/B_T(1-t,1)$ and the third by the change of variable $q=q(t,p)$. Hence, the induction step holds, and (43) is satisfied for all $k\in\{0,...,C_{dT}\}$. We obtain the desired expression by taking $t=0$ and $k=C_{dT}$.

If Assumption 2(ii) further holds, we obtain by Lemma D.1 that if $f=g_{\lambda,\mu}$, then $T(f;u)=g_{\lambda+1,\mu+u}$. Let $V_k(t,p;\lambda,\mu)$ and $V_k^*(t;\lambda,\mu)$ be defined as in the full dynamic pricing case. Then, by the same induction as above, we have, for all $k\in\{0,...,C_{dT}\}$,

$$V_k^*(t;\lambda,\mu)=\left[\frac{B_T(1-t,1)}{\mu}\right]^{\frac{1}{\varepsilon}}\alpha^i_{k,s}(\lambda), \tag{45}$$

where $\alpha^i_s(0,\lambda)=0$ for $\lambda>0$, and

$$\alpha^i_{k,s}(\lambda)=\max_{q>0}q\int_0^1\frac{\lambda}{(1+qs)^{\lambda+1}}\left[q^{-\frac{1}{\varepsilon}}+\left(\frac{1-s}{1+qs}\right)^{\frac{1}{\varepsilon}}\alpha^i_{k-1,s}(\lambda+1)\right]ds.$$

The result follows by taking $t=0$, $\mu=Y_{dT}^{-1}$, and $k=C_{dT}$, we obtain the desired expression.

Stopping-time pricing with M fares. As in the complete information case, let $V_k(0;t,p,m)$ (resp. $V_k(1;t,p,m,f)$) denote the optimal revenue at time $1-t$, with a current price p, a remaining capacity k, a remaining number of fares m and a density of f for ξ_{dT} (conditional on the current information) if the firm decides to keep the same price (resp. to change it). Then, as (33), we have:

$$
\begin{aligned}
V_k(0;t,p,m,f) &= \int_0^t f_{\tau_{1-t,p}}(s)\left[p + V_{k-1}^*(t-s,p,m,T(f;q(t,s,p)))\right]ds, \qquad (46)\\
V_k(1;t,p,m,f) &= \max_{p'>0}\int_0^t f_{\tau_{1-t,p'}}(s)\left[p' + V_{k-1}^*(t-s,p',m-1,T(f;q(t,s,p)))\right]ds,\\
V_k^*(t,p,m,f) &= \max_{d\in\{0,1\}} V_k(d;t,p,m,f),
\end{aligned}
$$

with the initial conditions $V_0^*(t,p,m,f)=0$. We prove by induction on k that for all $(k,m)\in\{0,...,C_{dT}\}\times\mathbb{N}$,

$$
V_k^*(t,p,m,f) = c_{k,m}(q(t,p),f)\left[B_T(1-t,1)\right]^{\frac{1}{\varepsilon}}, \qquad (47)
$$

where $c_{k,0}(q,f) = q^{-\frac{1}{\varepsilon}}\int \mathbb{E}[D(qz)\wedge k]f(z)dz$ and

$$
\begin{aligned}
c_{k,m}(q,f) = \max\Bigg\{ & q\int_0^1\int ze^{-qzu}f(z)dz\Big[q^{-1/\varepsilon} + c_{k-1,m\wedge(k-1)}(q(1-u),T(f;qu))\\
& (1-u)^{\frac{1}{\varepsilon}}\Big]du, \max_{q'>0} q'\int_0^1\int ze^{-q'zu}f(z)dz\Big[q'^{-1/\varepsilon}\\
& + c_{k-1,m-1}(q'(1-u),T(f;q'u))(1-u)^{\frac{1}{\varepsilon}}\Big]du\Bigg\}.
\end{aligned}
$$

The result holds for $k=0$ since $c_{0,m}=V_0^*(t,p,m,f)=0$. Suppose that it holds for $k-1\geq 0$ and all $m\leq k-1$ (recall that $V_k^*(t,p,m,f)=V_k^*(t,p,m\wedge k,f)$). If $m=0$, the price cannot be changed anymore, so $V_k^*(t,p,m)$ is simply the revenue with price p from $1-t$ to 1, and (34) holds.

If $m\geq 1$, we have, using (44) and (46) and the same change of variables as above, we obtain

$$
\begin{aligned}
& V_k(0;t,p,m,f) \\
=& \int_0^t \left[\int_0^\infty (\xi_a+\xi_b)p^{-\varepsilon}\partial_2 B_T(1-t,1-t+s)ze^{-q(t,s,p)z}f(z)dz\right] \\
& \qquad \left[p+c_{k-1,m\wedge(k-1)}(q(t-s,p),T(f;q(t,s,p)))[B_T(1-t+s,1)]^{\frac{1}{\varepsilon}}\right]ds \\
=& \int_0^1 \left[\int_0^\infty q(t,p)ze^{-q(t,p)uz}f(z)dz\right]\left[p+c_{k-1,m\wedge(k-1)}(q(t,p)(1-u),\right. \\
& \qquad \left. T(f;q(t,p)u))[B_T(1-t,1)]^{\frac{1}{\varepsilon}}(1-u)^{\frac{1}{\varepsilon}}\right]du \\
=& [B_T(1-t,1)]^{\frac{1}{\varepsilon}}q(t,p)\int_0^1 \left[\int_0^\infty ze^{-q(t,p)uz}f(z)dz\right] \\
& \quad \times \left[q(t,p)^{-1/\varepsilon}+c_{k-1,m\wedge(k-1)}(q(t,p)(1-u),T(f;q(t,p)u))(1-u)^{\frac{1}{\varepsilon}}\right]du.
\end{aligned}
$$

By the same reasoning and the change of variable $q=q(t,p)$,

$$
\begin{aligned}
V_k(1;t,p,m,f) = [B_T(1-t,1)]^{\frac{1}{\varepsilon}} \max_{q>0} q\int_0^1 \left[\int_0^\infty ze^{-qzu}f(z)dz\right] \\
\left[q^{-1/\varepsilon}+c_{k-1,m-1}(q(1-u),T(f;qu))(1-u)^{\frac{1}{\varepsilon}}\right]du.
\end{aligned}
$$

Then,

$$
\begin{aligned}
V_k^*(t,p,m,f) =& \max_{d\in\{0,1\}} V_k(d;t,p,m,f) \\
=& c_{k,m}(q(p),f)[B_T(1-t,1)]^{\frac{1}{\varepsilon}}.
\end{aligned}
$$

This concludes the induction step, proving that (47) holds for all $k\in\{0,...,C_{dT}\}$.

Stopping-time pricing with M increasing fares The proof follows by making the same changes as those made in the complete information set-up.

Intermediate-K stopping-time pricing The proof follows by making the same changes as those made in the complete information set-up.

E Method and results with time-varying elasticity

In this appendix, we suppose that travelers who purchase no later than the closing of fare class $S(=10)$ have price elasticity $\varepsilon_{\text{early}}$ and those who purchase afterwards have price elasticity $\varepsilon_{\text{late}}$. First, we describe the procedure of demand estimation, including the point estimates of $(\varepsilon_{\text{early}}, \varepsilon_{\text{late}})$, parameters of destination-train-specific effect β_0, distributional parameters $\lambda_0 = (\lambda_{a0}, \lambda_{b0})$, and set estimate of train-specific effect $g_0(.)$. Second, we derive the formula of the revenue with optimal uniform pricing under complete information in the presence of early and late purchasers and report its set estimate.

Demand estimation: $(\varepsilon_{\text{early}}, \varepsilon_{\text{late}})$, β_0, and λ_0. To accommodate $\varepsilon_{\text{early}}$ and $\varepsilon_{\text{late}}$, we need to modify Assumption 1:

Assumption 6 (Consumers' demand with $\varepsilon_{\text{early}}$ and $\varepsilon_{\text{late}}$) *For all T and $d \in \{a, b\}$, there exists $\varepsilon_{early}, \varepsilon_{late} > 1$ and a random process $b_T(.)$ on $[0, 1]$, continuous and satisfying $\min_{u \in [0,1]} b_T(u) > 0$ almost surely, such that conditional on ξ_{dT} and $b_T(.)$:*

1. *V_{dT} is a Poisson process with intensity $I_{dT}(t, p) = \xi_{dT} b_T(t) \varepsilon_{early} p^{-1-\varepsilon_{early}}$ for $(t, p) \in [0, t_S] \times [0, \infty)$ and $I_{dT}(t, p) = \xi_{dT} b_T(t) \varepsilon_{late} p^{-1-\varepsilon_{late}}$ for $(t, p) \in (t_S, 1] \times [0, \infty)$, where $t_S \in (0, 1)$ is the closing time of fare class S and is assumed to be a function of W_T. Without loss of generality, we let $\int_0^1 b_T(u) du = 1$.*

2. *V_{aT} and V_{bT} are independent.*

Under Assumptions 3 and 6, similarly to the first statement of Theorem 4.2, we point identify $(\varepsilon_{\text{early}}, \varepsilon_{\text{late}})$. If Assumptions 2 and 5 further hold, then we identify (β_0, λ_0). Table 10 summarizes the corresponding estimates.

Table 10: Binomial model of demand with $(\varepsilon_{\text{early}}, \varepsilon_{\text{late}})$

Price elasticity	
$\varepsilon_{\text{early}}$	4.55
$\varepsilon_{\text{late}}$	3.17
Destination(-train) specific effects	
Population (in M. inhabitants)	2.14
Regional capital	0.24
Travel time by train (in hours)	−1.94
Travel time by train, squared	0.33
Gamma distributions	
λ_{a0} (intermediate)	3.63
λ_{b0} (final)	2.62
Control for $X_d \times W_T$	Yes
R^2 of the reg. of $\ln(\xi_{bT}/\xi_{aT})$	0.501

Notes: The total number of trains is 2,909 and the total number of observations (fare classes × trains) is 21,988.

Demand estimation: $g_0(.)$. To partially identify $g_0(.)$, we still rely on consumers' rationality and weak optimality condition described in Section 4.3. Because $\varepsilon_{\text{early}} \neq \varepsilon_{\text{late}}$, the resulting moment inequalities differ from those under the assumption of constant price elasticity. The inequalities originating from consumers' rationality are modified as follows. First, for $k \leq S$ (recalling that t_S is a function of W_T),

$$\mathbb{E}\Big[\sum_{j=k}^{K} n_{djT} - C_T \wedge D\left(\exp\{X'_{dT}\beta_0\}\eta_{dT}\left[p_{dkT}^{-\varepsilon_{\text{early}}} g_0^{\text{early}}(W_T) + p_{d(S+1)T}^{-\varepsilon_{\text{late}}} g_0^{\text{late}}(W_T)\right]\right)\Big|W_T\Big] \leq 0, \quad (48)$$

and for $k > S$,

$$\mathbb{E}\Big[\sum_{j=k}^{K} n_{djT} - C_T \wedge D\left(\exp\{X'_{dT}\beta_0\}\eta_{dT} p_{dKT}^{-\varepsilon_{\text{late}}}\left[g_0^{\text{early}}(W_T) + g_0^{\text{late}}(W_T)\right]\right)\Big|W_T\Big] \leq 0, \quad (49)$$

where $g_0^{\text{early}}(W_T) = g_0(W_T)\int_0^{ts} b_T(t)dt$ and $g_0^{\text{late}}(W_T) = g_0(W_T)\int_{t_S}^{1} b_T(t)dt$. Note that the left-hand side of (48) is strictly decreasing in $\overline{g}(W_t) :=_{dkT}^{-\varepsilon_{\text{early}}} g_0^{\text{early}}(W_T) + p_{dST}^{-\varepsilon_{\text{late}}} g_0^{\text{late}}(W_T)$. Then, given d, $k \leq S$, and W_T, we obtain a lower bound on $\overline{g}(W_t)$. Similarly, we obtain a lower bound on $g_0^{\text{early}}(W_T) + g_0^{\text{late}}(W_T)$ from (49) given d, $k \leq S$, and W_T. As a result, for a given W_T, we obtain 24 linear constraints on $(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T))$.

Turning to the weak optimality condition, we have

$$\mathbb{E}[R_T(p_a, p_b)|W_T] = \max_{C_{aT}+C_{bT}=C_T} \Bigg\{ \sum_{d=a,b} p_d \int_0^\infty \mathbb{E}\Big[D\left(\exp\{X'_{dT}\beta_0\}\Big(p_d^{-\varepsilon_{\text{early}}} g_0^{\text{early}}(W_T) \right. \\ \left. + p_d^{-\varepsilon_{\text{late}}} g_0^{\text{late}}(W_T)\Big) z\right) \wedge C_{dT}\Big] g_{\lambda_{d0},1}(z) dz \Bigg\}.$$

Let $R(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T); \varepsilon_{\text{early}}, \varepsilon_{\text{late}}, \beta_0, \lambda_0) := \max_{k=1,\dots,K} \mathbb{E}[R_T(p_{akT}, p_{bkT})|W_T]$. Then, we obtain:

$$R(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T); \varepsilon_{\text{early}}, \varepsilon_{\text{late}}, \beta_0, \lambda_0) \leq \mathbb{E}\left[R_T^{\text{obs}}|W_t\right]. \tag{50}$$

Unlike the moment inequalities built on the consumers' rationality, the weak optimality condition does not deliver a linear constraint on $(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T))$ because of the maximization over k. The estimation of the identified set of the counterfactual revenue is complicated by this non-linearity. To circumvent this numerical challenge, we exploit a looser weak optimality inequality, namely

$$R(g_0^{\text{early}}, g_0^{\text{late}}; \varepsilon_{\text{early}}, \varepsilon_{\text{late}}, \beta_0, \lambda_0) \geq R(g_0^{\text{early}}, g_0^{\text{late}}; \varepsilon_{\text{early}}, \varepsilon_{\text{early}}, \beta_0, \lambda_0).$$

This inequality holds because $\varepsilon_{\text{early}} < \varepsilon_{\text{late}}$ and $p_{dkT} \geq 1$ for all T, $d \in \{a, b\}$ and $k \in \{1, ..., 12\}$. Then, (50) implies

$$R(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T); \varepsilon_{\text{early}}, \varepsilon_{\text{early}}, \beta_0, \lambda_0) \leq \mathbb{E}\left[R_T^{\text{obs}}|W_t\right].$$

In other words, the actual revenue management yields a higher revenue than the optimal uniform pricing strategy with a price chosen from the price grid and early purchasers's price elasticity. This looser inequality leads to an upper bound on $g_0^{\text{early}}(W_T) + g_0^{\text{late}}(W_T)$ and delivers a linear constraint.

To summarize, for a given W_T, consumers' rationality and the loosened version of the weak optimality condition imply linear constraints on $(g_0^{\text{early}}(W_T), g_0^{\text{late}}(W_T))$. The computation of the estimator of the identified set of counterfactual revenues then reduces to an optimization problem with linear constraints.

Counterfactual revenues. Given W_T and capacity C_{dT} for destination d, the expected revenue R_{udT}^c of the optimal uniform pricing strategy under complete information for destination d in train T is

$$R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}) = \int_0^\infty \max_{p_d>0} \Big\{ p_d \mathbb{E}\Big[D\Big(\exp\{X'_{dT}\beta_0\} \Big(p_d^{-\varepsilon_{\text{early}}} g_0^{\text{early}}(W_T) + p_d^{-\varepsilon_{\text{late}}} g_0^{\text{late}}(W_T) \Big) z \Big) \wedge C_{dT} \Big] \Big\} g_{\lambda_{d0},1}(z) dz,$$

and the expected optimal revenue at the train level is:

$$R^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_T) = \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}).$$

Denote by $\mathcal{G}$ the set of linear constraints on $(g_0^{\text{early}}, g_0^{\text{late}})$ derived in the previous section. Then, the lower bound of the set estimate of $R^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_T)$ is expressed as:

$$\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T) = \min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}).$$

Solving this program exactly is difficult. Rather, we obtain lower and upper bounds on $\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T$. First, note that using Jensen's inequality, we have

$$p_d^{-\varepsilon_{\text{early}}} g_0^{\text{early}} + p_d^{-\varepsilon_{\text{late}}} g_0^{\text{late}} \geq (g_0^{\text{early}} + g_0^{\text{late}}) p_d^{-\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})},$$

where $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}) = \frac{g_0^{\text{early}}\varepsilon_{\text{early}} + g_0^{\text{late}}\varepsilon_{\text{late}}}{g_0^{\text{early}} + g_0^{\text{late}}}$. Then,

$$R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}) \geq R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}),$$

with $R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT})$ the expected revenue of the optimal uniform pricing under constant price elasticity $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$. Then,

$$\begin{aligned}
&\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T) \\
&\geq \min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}).
\end{aligned}$$

Note that $R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT})$ is a function of $g_0^{\text{early}} + g_0^{\text{late}}$ and $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$. Moreover, for $d = a, b$ and any C_{dT}, it is increasing with respect to $g_0^{\text{early}} + g_0^{\text{late}}$ and decreasing with respect to $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$. Then

$$\max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT})$$

is also increasing with respect to $g_0^{\text{early}}+g_0^{\text{late}}$ and decreasing with respect to $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$. As a result, its minimization with respect to $(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}$ is achieved either when $g_0^{\text{early}} + g_0^{\text{late}}$ is minimized or $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$ is maximized. Denote the minimizer of $g_0^{\text{early}} + g_0^{\text{late}}$ by $\underline{g}_0$ and the maximizer of $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$ by $\overline{\varepsilon}$. Then,

$$
\begin{aligned}
&\min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}) \\
&= \min\Bigg\{ \min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}, g_0^{\text{early}}+g_0^{\text{late}}=\underline{g}_0} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), \varepsilon(g_0^{\text{early}}, g_0^{\text{late}}), g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}), \\
&\min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}, \varepsilon(g_0^{\text{early}}, g_0^{\text{late}})=\overline{\varepsilon}} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\overline{\varepsilon}, \overline{\varepsilon}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}) \Bigg\}.
\end{aligned}
$$

To perform the first minimization, it suffices to compute the upper bound of $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}})$ in $\mathcal{G}$ subject to $g_0^{\text{early}} + g_0^{\text{late}} = \underline{g}_0$, and compute the corresponding optimal revenue. Similarly, to perform the second minimization, it suffices to compute the lower bound of $g_0^{\text{early}} + g_0^{\text{late}}$ in $\mathcal{G}$ subject to $\varepsilon(g_0^{\text{early}}, g_0^{\text{late}}) = \overline{\varepsilon}$, and compute the corresponding optimal revenue with constant price elasticity $\overline{\varepsilon}$. The minimum of the two simulated revenues then bounds $\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T)$ from below. Second, note that

$$
\begin{aligned}
\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T) &\le \min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}, g_0^{\text{late}}=0} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, g_0^{\text{early}}, g_0^{\text{late}}, C_{dT}) \\
&= \min_{(g_0^{\text{early}}, g_0^{\text{late}}) \in \mathcal{G}, g_0^{\text{late}}=0} \max_{C_{aT}+C_{bT}=C_T} \sum_{d=a,b} R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{early}}, g_0^{\text{early}}, 0, C_{dT}).
\end{aligned}
$$

The revenue $R^c_{udT}(\varepsilon_{\text{early}}, \varepsilon_{\text{early}}, g_0^{\text{early}}, 0, C_{dT})$ is the expected revenue of the optimal uniform pricing under constant price elasticity $\varepsilon_{\text{early}}$ and only depends on g_0^{early}. Then it suffices to compute the lower bound of g_0^{early} in $\mathcal{G} \cap \{(g_0^{\text{early}}, 0) : g_0^{\text{early}} \in \mathbb{R}^+\}$ and simulate the corresponding optimal revenue with constant price elasticity $\varepsilon_{\text{early}}$ to bound $\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T)$ from above.

We estimate that the average of $\underline{R}^c_{uT}(\varepsilon_{\text{early}}, \varepsilon_{\text{late}}, C_T)$ among all trains lies between 12.81K€ and 13.39K€. This is quantitatively close to the lower bound of 13.23K€ obtained under the assumption of constant price elasticity and baseline specification, see Scenario u.4 in Table 4.