Demand Estimation in the Presence of Revenue Management

Xavier D’HAULTFOEUILLE, Philippe FÉVRIER et Lionel WILNER

Document de travail
Demand Estimation in the Presence of Revenue Management

Xavier D'HAUTFOEUILLE* - Philippe FÉVRIER** - Lionel WILNER***

OCTOBRE 2012

Les auteurs remercient Jean-Jacques BECKER, Éric DUBOIS, Johan HOMBERT, Claire LELARGE, Laurent LINNEMER, Corinne PROST et Michael VISSER pour leurs suggestions, et les participants aux séminaires INSEE-DEEEE et CREST-LEI pour leurs commentaires. Ils remercient également la SNCF et iDTGV pour leur avoir donné accès aux données.

* CREST-LMI
15, bd Gabriel Péri - 92245 MALAKOFF CEDEX

** CREST-LEI
15, bd Gabriel Péri - 92245 MALAKOFF CEDEX

*** Département des Études Économiques - Division Marchés et Entreprises
Timbre G230 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX
Demand Estimation in the Presence of Revenue Management

Abstract

Yield management has become a standard tool in several industries to increase the profits of firms facing demand uncertainty or consumers heterogeneity. But this technique also raises econometric problems in the estimation of demand models. Quantity-based management, in particular, is the source of both an endogeneity and a right-censoring problem. Disposing of macro data only and ignoring these issues leads to an aggregation bias. We develop a structural model of demand in the presence of quantity-based management. We show that the price elasticity is identified in this model provided that (i) we observe two subpopulations that face different prices but are not separated in the yield management policy, (ii) the highest prices and quantities sold at these prices are observed, (iii) the highest prices vary with time or across markets. We apply our method to the French railroad industry, using disaggregated data on trips between Paris and big cities on the period 2007-2009. Our estimates of the price-elasticity are consistent with a rather responsive demand, from 1.7 to 2 in economy class and from 1.3 to 1.5 in business class.

Keywords: revenue management, demand estimation, price-elasticity, railways transportation.

Estimation de la demande en présence de revenue management

Résumé

Le revenue management est devenu un outil standard de plusieurs industries, augmentant le profit des firmes faisant face à une demande incertaine ou des consommateurs hétérogènes. Mais cette technique soulève également des problèmes économétriques dans l’estimation de modèles de demande. Un revenue management basé sur les quantités, en particulier, est source à la fois d’endogénité et de censure à droite. Utiliser des données agrégées en ignorant ces problèmes conduit à un biais d’agrégation. Nous développons un modèle structurel de demande en présence d’un tel revenue management. Nous montrons que l’élasticité-prix est identifiée dès lors (i) qu’il existe deux sous-populations qui payent des prix différents mais ne sont pas distinguées par le revenue management, (ii) que les prix les plus élevés et quantités vendues correspondantes sont observés et (iii) que les prix les plus élevés varient dans le temps ou entre marchés. Nous appliquons notre méthode au transport ferroviaire en France, en utilisant des données détaillées sur les trajets entre Paris et plusieurs grandes villes sur la période 2007-2009. Nos estimations suggèrent une demande plutôt reactive, l’élasticité-prix se situant entre 1,7 et 2 en deuxième classe et entre 1,3 et 1,5 en première classe.

Mots-clés : revenue management, modèles structurels de demande, élasticité-prix, transport ferroviaire.

Classification JEL : C25, D12, R40.
1 Introduction

This paper proposes to estimate a structural model of demand in the presence of yield management. Talluri and van Ryzin (2005) define revenue management as a set of techniques devoted to the maximization of revenues. It typically involves demand-management decisions: structural decisions (related to the selling format, bundling issues, etc.), price decisions and quantity decisions. Price decisions refer to the (possibly dynamic) optimal pricing strategy. Quantity decisions may involve rationing schemes.

Yield management is only adopted by firms and industries selling perishable products and has been used by airlines, railways, hotel and rental car industries, telecommunications, and in the e-business. The e-commerce is a growing industry and firms resort increasingly to yield management techniques. We focus here on the railroad industry adopting quantity-based yield management with a pricing scheme of the following form: several fare classes are defined, in which the price is fixed. The question is to allocate a number of tickets (a booking limit or a quota) to each fare class.

Quantity-based management raises several issues in terms of demand estimation. First, the usual simultaneity issue in the formation of demand cannot be ignored: the price may vary because the opening and closing of fare classes adjust to current demand fluctuations. For this reason, endogeneity arises. Second, quotas are fixed by firms for every fare class. As a result, only the minimum of this quota and of the demand at that price can be obtained from data of purchases in every fare class. These quantities do not correspond to the actual demand at that price, which causes a right-censoring problem. Some of these issues have already been studied in the literature, especially the censoring issue (Swan (1990), Lee (1990) and Stefanescu (2012)).

Our contribution is (i) to provide a model of consumer behavior in the presence of quantity-based management; (ii) to explain under which conditions the price-elasticity of demand is identified in the data; (iii) to estimate our model of demand on prices and purchases available for every yield class of French trains resorting to such revenue management.

We propose to micro-found the demand thanks to a simple structural model that relies on individual decisions of purchases made by strategic and forward-looking consumers. The revenue management literature has devoted little attention to the micro-foundation of the demand. We assume that the flow of consumers’ arrivals on the market follows a Poisson process. This Poisson process is nonhomogeneous so that the instantaneous
probability of an arrival is allowed to vary over time. The revenue management literature has already resorted to such processes for consumers’ arrivals, at least since Beckmann and Bobkowski (1958). Lee (1990) and Gallego and van Ryzin (1994), among others, also assumed homogeneous or non-homogeneous Poisson processes. As a result, different consumers indexed by their time of arrival are on the market and may purchase or not, the outcome depending on their valuation. Depending on this date, consumers draw their valuation for the good from a Pareto distribution. Contrary to theoretical models in the yield management literature (Littlewood (1972), Brumelle and McGill (1993)), we do not impose that more price-sensitive consumers come first, which is a more general approach. Consumers who buy are those for which the valuation is higher than the corresponding fare. This last part, including a rational choice made by strategic and forward-looking agents, has received less attention from the literature, apart from the recent papers of Talluri and van Ryzin (2004, 2008), Ratliff et al. (2008) and Haensel et al. (2011).

We consider the case of the French railroad industry where such a quantity-based yield management is actually used. We estimate structural preferences for destinations and the price-elasticity of demand in this industry.

We obtain a rather high price-elasticity, which tends to suggest that the population is very responsive to price changes. The elasticity is found to be between 1.3 and 1.5 in the business class, and between 1.7 and 2 in the economy class. These results are high but still in the range of the estimates presented in the previous literature (see for instance Jevons et al. (2005), Oum et al. (1992), Rolle (1997), Wardman (1997, 2006) and Wardman et al. (2007) among others). On French data, Sauvant (2002) finds a price-elasticity of about 0.7.

In comparison with most of these papers in the transportation literature, the main advantage of this study stems from the very detailed level of information. Most studies cited above rely on macro data, and the authors cannot correct for the endogeneity bias. This bias is increased by the aggregation of data. Moreover, the aggregation itself is responsible for spurious price variations coming from changes in quantities sold and not in prices. On the contrary, using detailed data enables us to identify the price-effect and to control for the endogeneity problem as well.

Studying such revenue management practices may be of some interest in the construction of consumer price indices (CPI). Indeed, goods that are exposed to intertemporal variation (including promotions, increasing or decreasing price patterns) have already received attention from statisticians in charge of these indices. Facing dynamic price
patterns, consumers may react and try to buy in advance or later, which could imply to make further price listings and to weight every price accordingly, in relation with corresponding purchases – usually unobserved.

Section 2 is devoted to the yield management. We present our model in Section 3. We discuss identification issues in Section 4. In Section 5, we present an application to the railroad industry. Section 6 concludes.

2 Yield management practices

Historically, revenue management was first used by American airlines after the adoption of the Airline Deregulation Act in 1978. Companies were then allowed to change their prices, the schedules and the service without the approval of the regulator, namely the U.S. Civil Aviation Board. In Europe and notably in France, this practice has also been widely used over the last years. Such techniques are not restricted to airways: railways have also been resorting to them, as well as hotel and rental car industries, telecommunications, etc.

Firms may use several types of yield management techniques to price discriminate among their customers. The two main practices are price-based and quantity-based revenue management.

Price-based yield management is mainly used in the retailing industry and involves promotions, discounts and sales as instruments for dynamic pricing. “E-tailers” may resort to dynamic pricing and to adjustments based on real-time information.\(^1\) Finally, consumer-packages goods promotions in the retailing industry are a form of price-based management. Such practices are related to the stockpiling behavior of (at least some) consumers (see for instance Hendel and Nevo (2011)) and relate to intertemporal price discrimination models. They are typically devoted to durable or storable products, and may result in periods of regular prices followed by occasional sales.

In contrast, quantity-based management relies on the optimal allocation of capacity to different classes of demand that are called yield classes or fare classes. A unique price is assigned to each fare class. The firms’ strategy consists in defining “protection levels”, that is, amounts of capacity to reserve for a particular fare class (or set of fare

\(^1\)They also practice behavior-based price discrimination or customer recognition, the price being a function of purchase history. Such techniques may be based on e-tracking with the help of cookies (see Fudenberg and Villas-Boas (2006) for a survey of those practices).
classes). These protection levels can be viewed as class-specific quotas. Usually, quantity-based management implies that prices keep increasing over time, with the cheapest prices corresponding to the lowest fare class.\footnote{As a result, intertemporal price discrimination arises since according to the time of arrival on the market, the same good may be proposed at different prices.} Such a quantity-based management is typically applied to perishable goods; a seat in a plane or a train is perishable in the sense that it has no value after departure.

Note that if quantity-based management consists in rationing with a limiting-supply mechanism, price-based management may achieve the same objective by increasing prices, which reduces sales. As a result, these two forms of yield management are not necessarily conflicting; firms often use both of them. For instance, airlines may resort to advance-purchase discounts, which is a price-management practice, and moreover limit this number of discount seats, which is a quantity-based management.\footnote{The literature devoted to advance-purchase discounts includes for instance Shugan and Xie (2000), Shugan and Xie (2005), Möller and Watanabe (2010) and Nocke and Peitz (2007).}

Static models of quantity-based management have been developed over the last decades. They rely on several assumptions and determine the optimal quotas \textit{ex-ante} in every fare class. First, demands for different classes are independent. Second, and more importantly, demands are assumed to arrive in the order of increasing prices of the fare classes, which rules out overlapping intervals of demands. Third, these demands do not depend on the capacity controls, and especially on the availability of other classes. Fourth, group bookings are ruled out. Finally, firms are assumed to be risk-neutral, which is a rather reasonable assumption since such decisions repeat for numerous goods sold repeatedly over time. The most famous and earliest model is the model proposed by (Littlewood (1972)) also known as “Littlewood’s rule”. The extension of this rule to $K$ classes was provided by Brumelle and McGill (1993). An approximation of this rule to $K$ classes widely used in practice was provided by Belobaba (1989) and is called Expected Marginal Seat Revenue (EMSR).

Contrary to previous static models, dynamic models of pricing under revenue management (Gallego and van Ryzin (1994), Feng and Gallego (1995), Feng and Xiao (2000) and McAfee and te Velde (2008)) do not posit that the number of seats allocated in each fare class is fixed \textit{ex-ante}. They determine the optimal stopping times for every fare class, that is, the best moments for a monopoly to close the current yield class and to open the next one. In the rest of the paper, we focus on demand only. An interesting extension would consist in estimating both a supply and demand model, where the supply part
would come from theoretical times of fare classes closing/opening.

3 Model

In this section we present a structural model of demand in presence of a quantity-based management with $K$ yield classes (or fare classes). This model of demand relies on consumers’ individual decisions of purchase. In the following, we consider the railroad industry and goods will correspond to destinations; note that the same model could be applied to markets where such a quantity-based management is at stake.

First, we need Assumption 1 that simply recognizes the need of observing several subpopulations submitted to the same yield management rule.

**Assumption 1** (Goods). *In every yield class (or fare class), at least two different goods are proposed to the customers, and the yield management is the same regardless of the good.*

Note that either consumers may choose between distinct goods or distinct groups of consumers may be offered the same good. Typically, different goods may encompass different destinations but also different types of passengers. In the case of airways (resp. railways), Assumption 1 is likely to be verified when either two destinations are offered by the same flight (resp. train) or when different prices are offered to consumers according to some observable characteristics like age (student or elderly fares) or marital status (families fares) for instance. Indeed, it is reasonable to think that these different goods obey the same booking procedure, even though an optimal booking policy would consist in fixing quotas for every type. If several destinations are deserved by a train before its terminal, the protection levels (and the corresponding quotas) may be decided “uniformly” at the train level and not at the destination level, which is what Assumption 1 requires. In other words, quotas are supposed to apply to the whole yield class, which may be filled up either with customers for one or the other destination, as long as the total number of tickets does not exceed the quota.

In the rest of the paper, we consider two destinations $d \in (a, b)$ with $a$ being the terminal and $b$ the intermediate stop. Importantly, we do not allow consumers to choose their destination (see *infra*). By fare, we mean a couple $(p_{ak}, p_{bk})$ in the fare class $k$. To be more precise, the yield management puts the following constraint on the price pattern: for all trains $T$, $\exists t_1 = 0 < t_2 < \ldots < t_K < t_{K+1} = \bar{t}$ such that

$$p_d(t) = p_{dk} \quad \forall t \in [t_k; t_{k+1}), \forall k = 1, \ldots, K, \quad \forall d$$
such that once \( q_k \) tickets have been sold, the \( k^{th} \) fare class closes and the \( k + 1^{th} \) fare class opens at \( t = t_{k+1} \). Times \( t_k \) when fare classes close are thus endogenous.

Second, for the sake of flexibility, we assume that consumers’ arrivals \( N_{dT}(t) \) for the destination \( d \) of train \( T \) follow a non-homogeneous Poisson process with parameter \( \lambda_{dT}(t) \) on a fixed time interval \([0; T]\). \( t \) corresponds to the delay between the opening of the booking and the train’s departure. Arrivals refer to consumers showing up online or at a ticket office to purchase a ticket. Nothing guarantees that the distribution of the number of arrivals \( N_{dT}(t) \) is constant over time: in contrast, peaks of arrivals are likely to be observed when the booking starts, or at the end of the selling period and this is why we choose a nonhomogeneous Poisson process.\(^1\) The number of arrivals may also be increasing or decreasing over time. Non-homogeneous Poisson processes are characterized by the following relation:

**Assumption 2 (Consumers’ arrivals).**

\[
\lim_{dt \to 0} \frac{\Pr(N_{dT}(t + dt) - N_{dT}(t) = 1)}{dt} = \lambda_{dT}(t). \tag{1}
\]

Third, we assume that this population is constituted of heterogeneous consumers differing in their valuation for the trip and that these valuations are distributed according to a Pareto on \([v_{T}(t); +\infty)\):

**Assumption 3 (Heterogeneity of valuations).** For all destinations,

\[
F_{T}(v|t) = 1 - \left( \frac{v}{v_{T}(t)} \right)^{-\epsilon}. \tag{2}
\]

Assumption 3 imposes for consumers to have the same constant price-elasticity since \( \epsilon_{T}(p|t) = \frac{p f_{T}(p|t)}{1 - F_{T}(p|t)} = \epsilon \), regardless of their time of arrival. The distribution of valuations is assumed here to be independent from \( d \), which is rather well-suited to a population who does not choose the destination. More generally, we think that the choice of the destination is not too much an issue here: many travels are constrained (not only in the business class) and we make this assumption. In other words, we do not allow consumers to choose among destinations but assume rather that travelers have an obligation, either professional or personal, such that they must go to some given destination provided that they can afford it. Finally, for reasons that will be developed in Section 5, it would not be possible to estimate a significant proportion of distance-specific price-elasticities.

\(^1\)The Poisson assumption for consumers’ arrivals is standard in the revenue management literature, as documented by Talluri and van Ryzin (2004).
However, consumers’ expected valuation may evolve over time according to the lowest valuation \( v_T(t) \) of the sub-population arriving at \( t \). The behavior of \( v_T(t) \) with \( t \) gives some information about the correlation between valuations and time of arrival. In a context of intertemporal pricing, the knowledge of this parameter is crucial: indeed, an increasing pattern of prices over time is a best response to valuations increasing over time, which is the case if \( v_T(t) \) is nondecreasing. This corresponds to a situation where low valuation consumers arrive earlier in the market. Otherwise, a uniform price should be observed (see for instance Choné et al. (2012)).

Conditional on a time arrival \( t \geq t_K \), that is after the opening of the last class, the probability that a consumer buys equals:

\[
P(v \geq p_d(t)|t \geq t_K) = P(v \geq p_d(t)|p_d(t) = p_dK) = 1 - F(p_dK|t) = v_T(t) p_dK. \tag{3}
\]

At this point, another assumption is required: we need some separability between \( d \) and \( t \). In other words, the consumers’ arrival process must be separable between a train-specific component and a destination-specific component:

**Assumption 4 (Separability - 1).**

\[
\lambda_{dT}(t) = \xi_d \nu_T(t). \tag{4}
\]

Then the number of consumers arriving between \( t_k \) and \( t_{k+1} \) \( \forall k = 1, \ldots, K_T \) and who are ready to buy a ticket to destination \( d \) – that is, the instantaneous demand \( D_{dkT} \) for class \( k \) and for destination \( d \) in the train \( T \) – follows a Poisson distribution since it is the counting process of the Poisson process for arrivals:

\[
\lim_{dt \to 0} \frac{P(D_{dkT}(t + dt) - D_{dkT}(t) = 1)}{dt} = \xi_d \nu_T(t) v_T(t) p_d - \epsilon \forall d, \forall t \in [t_k; t_{k+1}). \tag{5}
\]

This expression corresponds to the product of the instantaneous probability of arrival (1) and of the probability of purchase conditional on arrival (3). If \( \theta = (\epsilon, \xi, \nu_T, v_T) \) is the vector of structural parameters, one obtains easily Proposition 1.

**Proposition 1.** Under Assumptions 1-4, the demand at date \( t \geq t_k \) in fare class \( k \) of train \( T \) for destination \( d \) follows a Poisson distribution:

\[
D_{dkT}(t)|t_k, p_{dkT}; \theta \sim P \left( \xi_d p^{-\epsilon}_{dkT} \int_{t_k}^{t} \nu_T(u) v_T(u) du \right) \forall d, \forall t \in [t_k; t_{k+1}). \tag{6}
\]

Noting the class-train specific effect \( A_{kT}(t) = \int_{t_k}^{t} \nu_T(u) v_T(u) du \), we have:

\[
D_{dkT}(t)|t_k, p_{dkT}; \theta \sim P \left( \xi_d p^{-\epsilon}_{dkT} A_{kT}(t) \right) \forall d, \forall t \in [t_k; t_{k+1}). \tag{7}
\]
4 Identification

The price-elasticity of demand is a crucial parameter of interest, both for economists and firms, since it impacts directly the optimal tarification. The econometrician interested in estimating this parameter has thus to wonder whether the price-sensitivity is identified in the data. In our setting, the identification of the demand’s response to prices raises two non standard issues, in addition to the well-known endogeneity problem in the estimation of demand models. First, because of quantity-based management and the class-specific quotas, the demand (dependent variable) is not exactly observed, but right-censored. Second, in such a setting, the identification of the price-elasticity of demand is far from being guaranteed since exogenous variation in prices is required to identify this parameter, but prices are constant within fare classes.

4.1 Endogeneity

In demand models, prices are suspected to be endogenous for at least two reasons. First, the price is expected to be positively correlated to unobserved determinants of demand, like quality for instance. Second, there might be a simultaneity bias due to the fact that prices are set contemporaneously to the formation of the demand.

In this quantity-based management setting, these two problems are likely to occur. The supply should react to a higher demand by closing quickly the first fare classes with a low price in order to charge a higher price. As a result, prices and idiosyncratic shocks of demand must be positively correlated. In airlines or railways, prices are higher in rush hours because the demand is higher. Even if the adjustment is not dynamic, the fixation of the quotas must somehow be based on past demand: contrary to the econometrician, the supply knows unobserved components of its demand and fixes quotas accordingly. As a result, a positive correlation between prices and unobserved components of demand is very likely and may be responsible for an upward (resp. downward) bias of the price-effect (resp. the price-elasticity).

In the rest of the paper, we will document empirical evidence in favor of the existence of such endogeneity and try to quantify the corresponding bias.

4.2 Right-censoring

As already mentioned, this form of quantity-based management is responsible for a right-censoring problem of the dependent variable: the true demand is partly observed.
To make this point clear, consider here the total demand as a function of prices $D(p)$ that would have been realized if a single price had been proposed (as opposed to the previous demand $D_k(t)$ in every fare class defined for $t \in [t_k; t_{k+1})$), and noting $n_j$ the number of tickets purchased in fare class $j$, we know only from the data that

$$D(p_k) \geq \sum_{j=k}^{K} n_j.$$  \hspace{1cm} (8)

Indeed, if consumers purchase at price $p_j > p_k$, we know for sure that they would have purchased also at price $p_k$. We cannot say more about $D(p_k)$ since the reverse is not true.

If $b_k$ is the booking limit (the quota), the econometrician only observes the number of tickets purchased $n_k$ in class $k$. Since $n_k = \min(D_k(t_{k+1}), b_k)$, this constitutes a lower bound of the demand $D_k(t_{k+1})$. Without further assumption on consumers’ behavior, the true demand cannot be recovered. However, it is worth mentioning that the demand in the highest fare class is exactly observed in the case where the capacity constraint is not binding, that is, when the booking is not full. To be more precise, in the highest fare class reached ($k = K$) and in the event when the train is not full we do know that:

$$D_K(t) = n_K.$$ \hspace{1cm} (9)

As a result, we avoid the right-censoring issue by focusing on this highest fare class reached, in cases where the train is not full. Note that such an estimation based on Equation (9) makes use neither of the data on full trains nor on data about lower fare classes, which is an important caveat. However, the use of the information contained in the other fare classes would necessarily cost some further assumption about the way the booking is filled-up.

### 4.3 Identification of the price-elasticity

Restricting our attention to class $K_T$ (there is some variation in the highest fare class reached), we will consider $t = \bar{t}$ for which $D_{dKT}(\bar{t}) = n_{dKT}$ and note the train effect in the last class reached $A_T = A_{KT}(\bar{t}) = \int_{t}^{\bar{t}} \nu_T(u)\psi_{\bar{T}}(u)du$, which gives:

$$D_{dKT}(\bar{t})|t; p_{dKT}; \theta \sim \mathcal{P}\left(\xi_d p_{dKT}^{-1} A_T\right).$$ \hspace{1cm} (10)

Hence $A_T$ plays the role of a train fixed-effect. We could think of estimating the fixed-effect Poisson model (10) by maximum likelihood. This would provide a consistent

\hspace{1cm}1In this subsection, subscripts $d$ and $T$ are removed for the sake of clarity.
estimate for \((\epsilon, \xi)\) (Lancaster (2000)), but not for the train-specific effect \(A_T\) since there are only as many observations per train as the number of distinct destinations. For the estimate to be consistent however, the number of observations for every train should tend to infinity. It could be tempting to parameterize the fixed-effect as some function of trains’ characteristics (schedule for instance). However, this would not solve the endogeneity issue coming from the potential correlation between the price and the fixed-effect \(A_T\), which could lead to underestimate the price-elasticity (in absolute) as soon as \(A_T\) is present.

As a result, to avoid the endogeneity issue, we get rid of the train fixed-effect and more generally of every component that is not destination-specific, by conditioning on \(n_{aKT} + n_{bKT}\), the sum of tickets sold in class \(K\), as soon as Assumption 1 is verified and at least destinations \(a\) and \(b\) are available for a route \(r\):

**Proposition 2.** Under Assumptions 1-4, one has:

\[
naKT | (naKT + nbKT = n) \sim B(n, \mu). \tag{11}
\]

**Proof.** We use the fact that a random variable following a Poisson distribution \((n_aK)\), conditional on the sum of several independent Poissons \((naKT\) and \(nbKT\), follows a binomial distribution with parameters \(n\) and \(\mu\), where \(\mu\) satisfies:

\[
\mu = \frac{A_T \xi_a p_{aKT}}{1 + e^{\Delta \log \xi - \epsilon \Delta \log p_{KT}}} \tag{12}
\]

and using the notation \(\Delta X = X_b - X_a\) for some covariate \(X\).

This method had already been used by Ridder et al. (1994). Note that Proposition 2 only holds for trains that are not full, for which by definition, the number of purchases observed in the last fare class \(K_T\) is inferior to the capacity of the train minus the sum of purchases in previous fare classes.

This model can be estimated simply by MLE.\(^1\) The likelihood of an observation \((n_{aKT} = k, n_{aKT} + n_{bKT} = n)\) has to be maximized with respect to \(\theta = (\epsilon, \xi)\) and writes:

\[
P(n_{aKT} = k | (n_{aKT} + n_{bKT} = n)) = \binom{k}{n} \mu^k (1 - \mu)^{(n-k)}. \tag{13}
\]

Since the fixed-effect disappears from the likelihood, there is no more correlation between price and any other component of the demand, which solves the endogeneity

---

\(^1\)Note that the estimation can also be done thanks to a simple logit model on reshaped data \(\forall i = 1, \ldots, n_{aKT} + n_{bKT}\) define \(Y_{iKT} = 1[i < n_{aKT}]\).
problem. However, the estimation based on Equations (12) and (13) depends on the ratios of prices between destinations $a$ and $b$, for a given train, in the last class reached $\Delta \log p_{KT}$. As a result, the identification of the price-effect requires both variation in the ratio $\frac{p_bKT}{p_aKT}$ (thus variation in $K$) and that this ratio be not 1: it cannot be obtained in the case where $p_aKT = p_bKT$. In other words, at least two points of the demand curve are needed in order to recover the shape of the demand curve, controlling for route effects. Note also the key role of Assumption 1: the variation required is necessarily obtained thanks to the existence of several subpopulations (here, going to destination $a$ or to destination $b$).

The identification of $\xi_d$ stems directly from variations of ratios of tickets sold within a train route and between destinations. In particular, the normalization $\xi_b = 1$ is required for every route. The coefficient $\xi_a$ gives information about the odds ratio of average demands for destinations $a$ and $b$ controlling for price effects.

To sum up, only the parameters $\epsilon$ for the price-sensitivity and $\xi_a$ accounting for relative preferences for destinations are identified. In Appendix A, we explain why it is not possible to recover more; in particular, even the observation of times of purchase combined with further separability assumptions between train and time effects would not enable us to identify the behavior of the demand over time.

Finally, interesting extensions would consist in positing a model on the outside option for railroad passengers (like taking the car, the plane, etc.) and to allow for a choice across destinations.

5 Application

5.1 Data

We use French data from iDTGV where a quantity-based management close to the one described in Section 3 is used.

We observe trains from October 2007 to February 2009. The destinations fall into six train routes $r$: (i) from Paris to Mulhouse through Strasbourg; (ii) from Paris to Perpignan through Nîmes or Montpellier; (iii) from Paris to Bayonne, Biarritz, Saint-Jean-de-Luz and Hendaye through Bordeaux; (iv) from Paris to Toulouse through Bordeaux; (v) from Paris to Toulon, Saint Raphaël, Cannes and Nice through Marseille; (vi) from Paris to Marseille through Aix-en-Provence or Avignon. A train $T$ has a schedule (day and hour of departure), a terminal and a direction (from Paris or to Paris). We
aggregate Nîmes and Montpellier into Nîmes/Montpellier; Aix-en-Provence and Avignon into Aix/Avignon; Toulon, Saint Raphaël, Cannes and Nice into Côte d’Azur; Bayonne, Biarritz, Saint-Jean-de-Luz and Hendaye into Côte basque, because for all those sets of close destinations, the fares are identical: for instance, it does not cost more to go to Montpellier than to Nîmes. As a result, once this aggregation has been made, all trains have two destinations: their final stop \( a \) and an intermediate stop \( b \). Some trains are scheduled during a “rush hour”, periods with peaks of demand such as Friday and Sunday nights, holidays and public holidays, etc. In this case, prices are set according to a price scale very similar to the one for regular hours, with the same fare classes, except that prices are uniformly higher.

A train systematically offers a business class and an economy class. We observe several fare classes within both the business class and the economy class. The number of fare classes differs in the business and in the economy class. For instance, prices run from 19 to 76.9 euros for a Paris-Bordeaux trip in economy class and in regular hours; they run from 39 to 104.9 euros for the same trip in rush hours.

To sum up, for each train \( T \), we observe some features of \( T \) in both business and economy class: its schedule, terminal (which will be confounded with the route \( r \)), direction (from or to Paris) and stops \( d \), as well as the sets of prices and tickets in each fare class \( \{(p_{dkT}, n_{dkT})\} \) \( \forall d \in (a,b) \) and \( \forall k \in [1;K_T] \) where \( K_T \) is the last class reached in train \( T \).\(^1\)

![Figure 1: Number of purchases per train in economy class (density)](image)

Figure 1 gives a nonparametric estimation of the density of trains’ filling-up. The

\(^1\)By “train”, we mean from now and by convenience either the business class or the economy class.
value of exact frequencies (on the y-axis) has been removed. In economy class, the event “full train” is frequent; in that case, the capacity constraint of the train is binding, and 350 seats are occupied. It also frequently occurs that in economy class trains are filled up at about 83% of their capacity.

19% of the total variation in the number of tickets sold comes from the heterogeneity between destinations, a major determinant of demand in the railroad industry. But fundamentally, most of the residual variance of demand is due to train effects. We believe that unobserved factors like weather conditions, local events, (national) holidays are widely responsible for the formation of demand. As a result, railroad companies respond to such a behavior with a different pricing: when they expect a high demand during some specific week-end for a given destination, they should fix smaller quotas in the lowest classes, as an optimal firm would do.

Table 1 shows how the demand for railroad transportation is divided into several services. The service from Paris to Marseille is the most popular. Interestingly, passengers for the intermediate stop Aix-en-Provence/Avignon take the train to Marseille rather than the train to Côte d’Azur: as a result, trains to Côte d’Azur are most of the time filled up with passengers going to Côte d’Azur, and less often with passengers for Aix-en-Provence or Avignon.¹ On the contrary, the service Paris-Mulhouse is used more by passengers for Strasbourg, the intermediate stop. The same phenomenon occurs on the train route to Perpignan through Nîmes-Montpellier. We do not see major differences between the business and the economy classes in this respect, except in the case of Bordeaux, for which the demand in business is relatively lower.

<table>
<thead>
<tr>
<th>From</th>
<th>Total</th>
<th>Terminal</th>
<th>Stop</th>
<th>Total</th>
<th>Terminal</th>
<th>Stop</th>
<th>Total</th>
<th>Terminal</th>
<th>Stop</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
</tr>
<tr>
<td>Côte basque</td>
<td>80</td>
<td>53</td>
<td>22</td>
<td>27</td>
<td>16</td>
<td>248</td>
<td>141</td>
<td>70</td>
<td>107</td>
<td>53</td>
</tr>
<tr>
<td>Bordeaux</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marseille</td>
<td>159</td>
<td>93</td>
<td>18</td>
<td>66</td>
<td>14</td>
<td>307</td>
<td>187</td>
<td>27</td>
<td>120</td>
<td>25</td>
</tr>
<tr>
<td>Aix/Avignon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mulhouse</td>
<td>74</td>
<td>22</td>
<td>8</td>
<td>52</td>
<td>14</td>
<td>195</td>
<td>49</td>
<td>16</td>
<td>146</td>
<td>30</td>
</tr>
<tr>
<td>Strasbourg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Côte d’Azur</td>
<td>141</td>
<td>113</td>
<td>28</td>
<td>28</td>
<td>16</td>
<td>254</td>
<td>206</td>
<td>51</td>
<td>48</td>
<td>24</td>
</tr>
<tr>
<td>Aix/Avignon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perpignan</td>
<td>125</td>
<td>49</td>
<td>22</td>
<td>76</td>
<td>14</td>
<td>286</td>
<td>78</td>
<td>28</td>
<td>208</td>
<td>40</td>
</tr>
<tr>
<td>Nîmes/Montpellier</td>
<td>84</td>
<td>48</td>
<td>13</td>
<td>38</td>
<td>13</td>
<td>284</td>
<td>174</td>
<td>38</td>
<td>110</td>
<td>40</td>
</tr>
<tr>
<td>Toulouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bordeaux</td>
<td>86</td>
<td>48</td>
<td>13</td>
<td>38</td>
<td>13</td>
<td>284</td>
<td>174</td>
<td>38</td>
<td>110</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 1: Number of purchases per train by destinations

Prices and fare classes are fixed at a national level but the quotas of fare classes may vary for each train, depending on the level of the demand. As already explained,

¹Both trains run on the same route, thus none of them is faster than the other.
quantity-based revenue management consists in allocating booking limits to every fare class. Importantly, these quotas are determined for each class at the train level and thus apply to the sum of tickets over all destinations. For instance, quotas are set for the fare classes of the train Paris-Côte Basque on March, 30th 2008, and not for every trip Paris-Bordeaux, Paris-Biarritz, Paris-Bayonne, Paris-Hendaye, Bordeaux-Hendaye, etc. As a result, at least two options (destinations) follow the same yield management rule and thus Assumption 1 is verified.

Note that quotas may or not be reached. For instance, if a train goes from Paris to Nice by Marseille, the quota of class $k$ depends not only on the demand for Marseille, but also on the demand for Nice. A quota of 100 tickets may be reached with 80 tickets for Marseille and 20 for Nice, and vice-versa. Quotas may not be reached if the total demand for Marseille and Nice does not exceed 100 people. In the case where the quota for class $k$ has been reached, class $k + 1$ opens to booking and tickets are available at the new price $p_{k+1}$. The process goes on until the train leaves: either the train is not filled up, or there is an excess demand, in which case the train’s capacity constraint is binding. The so-called duplex trains – with two stages – may carry up to 545 passengers, 197 in the business class and 348 in the economy class. In practice, the capacity constraint is sometimes reached in economy, but almost never in business.

![Figure 2: Variation in relative prices between terminal and intermediate stop for the last fare class](image)

It is helpful to have a rough idea of the “optimality” of yield management. To that purpose, we compute the correlation between the number of tickets sold in low fare classes...
and the number of tickets sold in high classes. On Paris-Toulouse and Paris-Hendaye for instance, this correlation amounts to $-0.69$ when we compare purchases in the first three classes with those in the last three classes. This negative sign is consistent with a correct revenue management practice. Indeed, to maximize revenues when the demand is high, and thus when high classes are reached, it is necessary to close low classes quickly. The strongly negative correlation that we observe indicates that it is actually the kind of strategy that is adopted here by railroad companies. In other words, when a large number of tickets with a high price has been sold, it generally means that few tickets were sold with a low price. This statistics is also consistent with the presence of endogeneity and especially a positive correlation between prices and unobserved components of demand. Indeed, a correct management of fare classes moves prices in the same direction as unobserved components of demand (higher prices when the demand is higher).

Finally, Figure 2 provides information on the variation of $\Delta \log p_K$, that is, the only source of variation that enables the econometrician to identify the price-elasticity. On both the business and the economy classes, this covariate exhibits enough variation to be able to infer the price-effect.

Note however that for all trains to Marseille-Aix/Avignon and to Mulhouse-Strasbourg, we have no variation in the relative prices for destinations A and B (in both cases, prices are always identical for destinations A and B: $\log p_{aK} = \log p_{bK}$). As a result, it would not be possible to estimate a separate model for each of the six routes and therefore to assume that price-sensitivity is route-specific, since at least two of these route-dependent price elasticities would not be identified in the data. This is the reason why we estimate our model under the assumption that the price-elasticity is the same for all routes.

5.2 Results

Table 2 provides the results from the estimation of the binomial model of demand summarized by Equations (11) and (12). The estimated price-elasticity $\hat{\epsilon}$ in economy class is close to 2 while it is about 1.4 in business class. The estimations of tastes for destinations have the following interpretation: they suggest that the demand for the terminal Côte d’Azur is much higher than the demand for the intermediate stop Aix-en-Provence/Avignon. On the contrary, the demand for Mulhouse or Perpignan, two terminals, is lower than the demand for the intermediate stops, respectively Strasbourg and Nîmes/Montpellier. In economy class, these ratios can be ordered as follows from the lowest to the highest: Perpignan/Nîmes-Montpellier, Mulhouse/Strasbourg, Côte Basque/Bordeaux, Marseille/Aix-Avignon, Toulouse/Bordeaux and Côte d’Azur/Aix-

17
Avignon.

It may be interesting to add some further structure on preferences for destinations \( \log \xi \) at the cost of specifying a multiplicative model. For instance, \( \log \xi \) could be some function of the distance and the time of the trip, as well as of the arrival city’s population measured in millions of inhabitants. We form therefore the ratio of the distance (resp. the time) between Paris and the intermediate stop on the one hand, and the distance (resp. the time) between Paris and the last stop on the other hand. We proceed similarly for the corresponding ratios of population. In the same vein, we have a proxy for the intensity of the substitution with airlines by considering the daily number of Air France flights for a given destination.

Table 2: Binomial model of demand
Equation (11)

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log (\text{price}) )</td>
<td>-1.441***</td>
<td>-1.994***</td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>Côte Basque</td>
<td>-0.066***</td>
<td>-0.610***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Marseille</td>
<td>0.248***</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Mulhouse</td>
<td>-1.085***</td>
<td>-1.444***</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Côte d’Azur</td>
<td>1.145***</td>
<td>0.844***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Perpignan</td>
<td>-0.663***</td>
<td>-1.457***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Toulouse</td>
<td>0.185***</td>
<td>0.511***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>( N )</td>
<td>4,413</td>
<td>5,595</td>
</tr>
</tbody>
</table>

\( N \): number of (non full) trains with Paris being either the departure or the terminal.

In business class, the average odds ratio of demands for Mulhouse (terminal) wrt Strasbourg (intermediate stop) is \( e^{-1.085} \approx 0.34 \); the price-elasticity is 1.441.

Assumption 5 (Structure).

\[
\log \xi = \alpha_0 + \alpha_1 \log(\text{distance}) + \alpha_2 \log(\text{time}) + \alpha_3 \log(\text{population}) + \alpha_4 \log(\text{airline}). \tag{14}
\]

Table 3 provides results of the estimation under Assumption 5. The estimated price-elasticities are still higher, between 1.61 and 2.66. The effects of distance, travel time and

\(^1\)the departure being Paris
population all have the same expected and positive sign, which indicates that demand is higher *ceteris paribus* for a destination that is further away, longer to serve and larger. The effect of distance may account for the substitution with the car: as distance increases, choosing the car becomes less a plausible outside option. Substitution with airlines should go the opposite way, since as the distance increases, choosing the airplane is more likely.

Table 3: Binomial model - Equation (11) with Assumption 5

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(price)</td>
<td>$-1.607^{***}$</td>
<td>$-2.663^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.962***</td>
<td>0.827***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Distance</td>
<td>1.618***</td>
<td>3.472***</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.192)</td>
</tr>
<tr>
<td>Time</td>
<td>1.744***</td>
<td>1.094***</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Population</td>
<td>0.690***</td>
<td>0.662***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Airline</td>
<td>$-0.173^{***}$</td>
<td>0.377***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>$N$</td>
<td>4,413</td>
<td>5,595</td>
</tr>
</tbody>
</table>

$N$: number of (non full) trains with Paris being either the departure or the terminal. 
In business class, the price-elasticity is 1.607. 
Distance in kilometers, time in minutes, population in millions inhabitants, airline in number of daily flights.

To check how much is gained with a nonparametric approach for the train fixed-effect, we also estimate a model based on Equation (10) where the effect $A_T$ is some function of the time of departure (month $m$ and day $d$) and of the type of period $p$ (normal or rush hour):

$$D_{dKT} \sim \mathcal{P}(A_0 \xi_d p_{dKT}^{-1} e^{\delta_m 1_m + \delta_d 1_d + \delta_p 1_p}). \tag{15}$$

However, as explained above, this approach is valid only if the endogeneity issue can be ignored and provides a consistent estimation only if there is no correlation between the price and unobserved demand effects once month/day/period effects have been controlled for. Results are shown in Table 4. They are rather conform to the intuition. The demand is higher in December, July, August and May, i.e. during holidays or public holidays. *Ceteris paribus*, the demand is higher during rush hour and week-ends. Marseille is the most frequented route, followed by Côte d’Azur, Perpignan (maybe because of the intermediate stop Nîmes/Montpellier), Toulouse, Côte basque and Mulhouse. Finally,
Table 4: Poisson model of demand - Equation (15)

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.335***</td>
<td>10.003***</td>
</tr>
<tr>
<td>log (price)</td>
<td>-1.496***</td>
<td>-1.708***</td>
</tr>
<tr>
<td>Côte basque</td>
<td>-0.144***</td>
<td>-0.313***</td>
</tr>
<tr>
<td>Marseille</td>
<td>+0.984***</td>
<td>+0.935***</td>
</tr>
<tr>
<td>Mulhouse</td>
<td>-0.326***</td>
<td>-0.344***</td>
</tr>
<tr>
<td>Côte d’Azur</td>
<td>+0.753***</td>
<td>+0.831***</td>
</tr>
<tr>
<td>Perpignan</td>
<td>+0.645***</td>
<td>+0.789***</td>
</tr>
<tr>
<td>Toulouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal</td>
<td>-0.479***</td>
<td>-0.482***</td>
</tr>
<tr>
<td>rush hour</td>
<td>ref.</td>
<td>ref.</td>
</tr>
<tr>
<td>Monday</td>
<td>-0.343***</td>
<td>-0.444***</td>
</tr>
<tr>
<td>Tuesday</td>
<td>-0.156***</td>
<td>-0.317***</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-0.074***</td>
<td>-0.222***</td>
</tr>
<tr>
<td>Thursday</td>
<td>0.021</td>
<td>-0.033***</td>
</tr>
<tr>
<td>Friday</td>
<td>-0.086***</td>
<td>-0.169***</td>
</tr>
<tr>
<td>Saturday</td>
<td>-0.146***</td>
<td>0.059***</td>
</tr>
<tr>
<td>Sunday</td>
<td>ref.</td>
<td>ref.</td>
</tr>
<tr>
<td>January</td>
<td>-0.224***</td>
<td>-0.455***</td>
</tr>
<tr>
<td>February</td>
<td>-0.062***</td>
<td>-0.374***</td>
</tr>
<tr>
<td>March</td>
<td>-0.294***</td>
<td>-0.476***</td>
</tr>
<tr>
<td>April</td>
<td>0.052**</td>
<td>-0.163***</td>
</tr>
<tr>
<td>May</td>
<td>0.019</td>
<td>0.011</td>
</tr>
<tr>
<td>June</td>
<td>0.061***</td>
<td>-0.030***</td>
</tr>
<tr>
<td>July</td>
<td>0.094***</td>
<td>0.023***</td>
</tr>
<tr>
<td>August</td>
<td>0.180***</td>
<td>-0.001***</td>
</tr>
<tr>
<td>September</td>
<td>-0.259***</td>
<td>-0.174***</td>
</tr>
<tr>
<td>October</td>
<td>-0.244***</td>
<td>-0.558***</td>
</tr>
<tr>
<td>November</td>
<td>-0.195***</td>
<td>-0.435***</td>
</tr>
<tr>
<td>December</td>
<td>ref.</td>
<td>ref.</td>
</tr>
<tr>
<td>N</td>
<td>8,840</td>
<td>11,112</td>
</tr>
</tbody>
</table>

N: destination-(non full) trains.

The average demand in business class a Sunday of December on rush hour is $e^{8.335} p^{-1.496}$. 

20
all these effects encompassed by $\delta$ are found to be very similar in both the business and the economy classes, and the average demand is higher in economy. According to these estimates, the price-elasticity tends to be about 1.50 (resp. 1.71) in business (resp. economy) class. In comparison with the previous estimation, especially in economy where the estimated elasticity was close to 2, this lower price-elasticity in absolute is consistent with the presence of some endogeneity that introduces a downward bias towards zero. The seemingly small magnitude of this bias, in particular in the business class, would tend to suggest that controlling for month, day and period (rush hours/regular schedule) effects treats much of the endogeneity problem.

For further robustness checks, it can be tempting to modify slightly Equation (15) and to replace $\epsilon$ with $\epsilon_{dT}$ in order to recover a distribution of price-elasticities across destinations, months, days or types of period. This experiment is provided by Table 5. The more constrained the demand is, as is the case during holidays, public holidays or week-ends, the less price-sensitive it is, which is rather intuitive. Overall, the price-sensitivity is low on Sundays, rush hours and in December. Moreover, the demand is almost always more price-sensitive in economy than in business class. Finally, the population of customers going to Toulouse is rather less price-sensitive than the one going to Marseille or Mulhouse for instance, which could be explained by the fact that the trip to Toulouse is long (about 5 hours, because of the absence of a high-speed rails) and that the population resorting to the train is very specific and has no other outside option.

Finally, to quantify more precisely the endogeneity bias in a reduced-form approach, it is interesting to perform the following regressions which correspond to a log-log estimation of the demand:

$$\log D(p_{dKT}) = -\epsilon \log p_{dKT} + \delta_T + \zeta_r + \eta_{dKT}. \quad (16)$$

and its corresponding first-difference equation:

$$\Delta \log D(p_{KT}) = -\epsilon \Delta \log p_{KT} + \Delta \zeta_r + \Delta \eta_{KT}. \quad (17)$$

Note that this reduced-form approach can also be seen as a robustness check with respect to the previous structural model. Equation (16) measures a reduced-form price-elasticity where the train-effect has been parameterized while Equation (17) is the corresponding estimation in first difference (model within). For the reasons exposed in Section 4.1, Equation (16) suffers from an endogeneity bias because the train component $\delta_T$ is correlated to $p_{dKT}$ while Equation (17) should not suffer from such a bias.\(^1\)

\(^1\)The analogue in the structural approach was the comparison between Equations (10) and (15).
Table 5: Robustness checks on the price-elasticity $\epsilon$

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$ Côte basque</td>
<td>1.445*** (0.039)</td>
<td>1.188*** (0.027)</td>
</tr>
<tr>
<td>$\epsilon$ Marseille</td>
<td>1.738*** (0.028)</td>
<td>2.120*** (0.017)</td>
</tr>
<tr>
<td>$\epsilon$ Mulhouse</td>
<td>1.743*** (0.052)</td>
<td>1.441*** (0.029)</td>
</tr>
<tr>
<td>$\epsilon$ Côte d’Azur</td>
<td>1.297*** (0.029)</td>
<td>1.564*** (0.016)</td>
</tr>
<tr>
<td>$\epsilon$ Perpignan</td>
<td>1.569*** (0.049)</td>
<td>1.979*** (0.017)</td>
</tr>
<tr>
<td>$\epsilon$ Toulouse</td>
<td>1.068*** (0.054)</td>
<td>0.962*** (0.043)</td>
</tr>
</tbody>
</table>

Panels separated by two lines correspond to different regressions.

Benchmark: Table 4/Equation (15).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8,840</td>
</tr>
</tbody>
</table>

$N$: destination-(non full) trains.

Round effect in business class, 1,706 in economy class.
Table 6: Reduced-form evidence on the endogeneity bias - Equations (16) and (17)

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>level</td>
<td>first difference</td>
</tr>
<tr>
<td>log (price)</td>
<td>$-1.013^{***}$</td>
<td>$(0.051)$</td>
</tr>
<tr>
<td></td>
<td>$-1.300^{***}$</td>
<td>$(0.058)$</td>
</tr>
<tr>
<td>model within</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>route effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>month effects</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>day effects</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>period effects</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.10</td>
<td>0.32</td>
</tr>
<tr>
<td>$N$</td>
<td>7,210</td>
<td>2,790</td>
</tr>
</tbody>
</table>

Table 6 shows the results: one can see that the endogeneity bias may be responsible of a downward attenuation of about $-0.29$ to $-0.43$ of the price-elasticity. Moreover, this reduced-form estimation leads to estimates that are close to the previous ones, such that a range of price-elasticities can be $[1.3; 1.5]$ (resp. $[1.7; 2]$) in business (resp. in economy) class. Note that Equation (17) provides a reduced-form estimation of the price-elasticity of demand that is perhaps more reliable in the presence of yield management.

5.3 Aggregating data

The estimation of demand (and thus of the price-elasticity) at a macro level is somehow problematic because aggregation leads to an attenuation bias that biases estimates towards zero. To illustrate, we aggregate our micro data and estimate 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. Formally, let $D_{dT}$ be the total demand for destination $d$ in the train $T$: $D_{dT} = \sum_{k=1}^{K_T} n_{dkT}$ and the average price $\bar{p}_{dT}$ is given by:

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

We could get rid of the train fixed-effect by considering the regression in differences as soon as at least two destinations $a$ and $b$ are available (Assumption 1):

$$\Delta \log D_T = -\epsilon \Delta \log \bar{p}_T + \Delta \log \xi_r + \delta_t + \Delta \eta_T,$$  

(18)
where $\log(\xi_{dr})$ accounts for a destination-specific component if $d \in (a, b)$ and $r$ is some train route. Results are given in Table 7. The estimated price-elasticity of 0.67 is reminding about Sauvant (2002) who found a value close to 0.7. However, the demand would be more responsive to prices in business than in economy, which is rather counterintuitive.

Similarly, we can aggregate data at the train level:

$$D_T = \sum_d \sum_{k=1}^{K_T} n_{dkT},$$

define an average price:

$$\bar{p}_T = \frac{\sum_d \sum_{k=1}^{K_T} n_{dkT} \bar{p}_{dkT}}{\sum_d \sum_{k=1}^{K_T} n_{dk}},$$

and perform identical regressions by controlling for the day of departure (model train):

$$\log D_T = -\epsilon \log \bar{p}_T + \delta_t + \eta_T, \quad (19)$$

Finally, the most aggregated approach (a time series approach) consists in aggregating these demands at a weekly or monthly level, either by train route or at the national level. Results show the existence of the attenuation bias. We conclude that the time series approach measures rather a covariation of quantities with prices than the microeconomic parameter of interest, the price-elasticity.

Table 7: Estimated price-elasticity from aggregated data - Equations (18) and (19)

<table>
<thead>
<tr>
<th></th>
<th>Business class</th>
<th>Economy class</th>
</tr>
</thead>
<tbody>
<tr>
<td>model within</td>
<td>1.61***</td>
<td>0.67***</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>model train</td>
<td>0.22***</td>
<td>0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>model week (by route)</td>
<td>0.56***</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>model month (by route)</td>
<td>1.41**</td>
<td>0.70**</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.31)</td>
</tr>
<tr>
<td>model week (France)</td>
<td>0.76***</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>model month (France)</td>
<td>1.64***</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(0.40)</td>
</tr>
</tbody>
</table>

The opposite of the log-price coefficient is reported for each model.

Refer to the text for an explanation of each model.
5.4 Discussion

The literature devoted to the transportation industry provides several estimations of the price-elasticity. Most studies rely on aggregated data (see for instance the meta-analysis by Jevons et al. (2005) and several papers like Wardman (1997), Wardman (2006) and Wardman et al. (2007)). The elasticities that were obtained by these authors, in the meta-analysis and in their econometric study as well, belong to [1.3; 2.2]. On German data, using a market-share approach, Ivaldi and Vibes (2005) find an elasticity close to the lowest part of this range, about 1.3. On Swiss data, Rolle (1997) has a price-elasticity of 1.5. His paper is focused on rather short travels and he is noticing that the elasticity tends to increase with distance, which could explain our higher elasticities in the French case since distances are longer. Other papers point out a lower elasticity, like the survey by Oum et al. (1992) presenting elasticities varying between 0.1 and 1.5. Similarly, Sauvant (2002) finds a price-elasticity of 0.7 on French data.

The difference between our results and those from studies relying on aggregated data comes precisely from the fact that we dispose of very disaggregated data. We already mentioned in Section 5.3 that the approach based on aggregated data leads to bias towards zero the estimated price-elasticities. In presence of such quantity-based management, disposing of very fine data at the fare class level seems necessary to identify the price-sensitivity.

Finally, our results may even provide lower bounds for the price-elasticity of demand. Indeed, the estimation relies on tickets bought in the highest fare class, which must correspond to less price sensitive consumers. Otherwise, we argued that it is hard to rationalize an increasing pattern over time. Though our model assumes the price-elasticity is constant, it is likely that more eager consumers come first. As a result, the price-elasticity should be lower in the highest class and higher in the lowest classes. More generally, if the true demand were observed in each of the $K$ fare classes, we would be able to estimate $\forall k = 1, \ldots, K$ the collection of price-elasticities $\epsilon_k$. In this paper, we provide an estimation of $\epsilon_K$ only, and perhaps the lowest – though already high.

6 Conclusion

The use of very fine data enables us to estimate the price-elasticity in the French railroad industry and structural parameters of a model of demand in presence of quantity-based revenue management. We explained how this form of yield management creates two econometric issues: a standard endogeneity problem and a non-standard right-censoring
problem in demand models. In particular, disposing of aggregated data in presence of such a pricing seems not enough to recover the true price effect. We provided a simple solution to overcome both problems and estimated a structural binomial model of demand. Railroad passengers are found to be rather responsive to prices since this price-elasticity varies from $1.3 - 1.5$ in business class to $1.7 - 2$ in economy class.
References


Appendix: times of purchase

In our model, it is not possible to disentangle what is due to a higher flow of arrivals, namely a strong potential demand, from what is due to large valuations. Generally speaking, a high demand can come either from a large market size and low individual propensions of purchase, or from a small market size and high individual propensions of purchase. In other words, $\nu_T(t)$ and $\nu_T(t)$ cannot be simultaneously identified: only the product $\nu_T(\cdot)\nu_T(\cdot)$ can be recovered.

Moreover, we face an incidental parameters problem since $\nu_T(\cdot)$ depends on the train. To be more precise, we need some further separability assumption between time and train effects in the product of arrivals’ flow and consumers’ minimum valuations (to get rid of the incidental parameters problem):

**Assumption 6** (Separability - 2).

$$\nu_T(\cdot)\nu_T(\cdot) = f(T) h(\cdot).$$

Furthermore, we also need to observe the individual times of purchase. In that case, one could recover the function of time $h(\cdot)$ that gives information about how the demand behaves over time. This is a crucial issue since it indicates among others, whether the demand is more eager at the beginning or not. Such a behavior is encompassed by the effect $A_{KT}(t) = \int_{t_K}^t \nu_T(u)\nu_T(u)du = f(T)\int_{t_K}^t h(u)du = f(T)H(t)$.

The behavior of $h(\cdot)$ is interesting from a microeconomic point of view. It is hard to rationalize the optimality of this quantity-based yield management practice if $h(\cdot)$ were some decreasing function over time; uniform pricing would then be optimal. On the contrary, when $h(\cdot)$ is nondecreasing over time, then some nondecreasing price pattern is optimal.

Since $\forall t \geq t_K$ one has $D_{dkT}(t)|t_K \sim \mathcal{P}\left(f(T)\xi dp_{dkT}^{-}\xi H(t)\right)$, one can use the same method as previously:

$$D_{dkT}(t)|D_{dkT}(t) + D_{dkT}(t') \sim \mathcal{B} \left(D_{dkT}(t) + D_{dkT}(t'), \frac{1}{1 + \frac{H(t')}{H(t)}}\right).$$

**Proposition 3.** If times of purchase are observed, and under Assumption 1-6, the function $h(\cdot)$ is identified nonparametrically up to a scale over $[\inf_T t_{K_T}; T]$. 

30
Proof. From the data and

\[
D_{d_kT}(t)\left|D_{d_kT}(t) + D_{d_kT}(t')\right| \sim B \left( D_{d_kT}(t) + D_{d_kT}(t') \frac{1}{1 + \frac{H(t')}{H(t)}} \right),
\]

the ratio \(\frac{H(t')}{H(t)}\) is identified. By fixing \(t\) and making \(t'\) vary, one identifies \(H(t')\) up to a scale over \([\inf_T t_{K_T} ; \bar{T}]\). As a result, its derivative \(h(\cdot)\) is identified up to a scale. \(\square\)

Note yet that some model of supply derived from the quantity-based management would add further constraints on the model, and this supplementary structure might help in achieving some functional identification of these two forces.
Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984-2002

G2009/07 S. QUANTIN - S. RASPILLER - S. SERRAVALLE
Commerce intra-groupe, fiscalité et prix de transferts : une analyse sur données françaises et étrangères

G2009/08 M. CLERC - V. MARCUS
Élasticités-prix des consommations énergétiques des ménages

G2009/09 G. LALANNE - E. POULIQUEN - O. SIMON
Prix de pétrole et croissance potentielle à long terme

G2009/10 D. BLANCHET - J. J. CAUVIN - V. MARCUS
Adjusted net savings and other approaches to sustainability: some theoretical background

G2009/11 B. BELLAMY - G. GONZALEZ - M. FESSEAU - S. LE LAIDER - É. RAYNAUD
Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003

G2009/12 J. BARDAJ - F. TALLET
Detecting Economic Regimes in France: a Quantitative Markov-Switching Indicator Using Mixed Frequency Data

G2009/13 R. AEBERTHARDT - D. FOGUIÈRE - R. RATHLEOT
Démocritus Discrimination à l'emploi : comment exploiter les procédures de testing ?

G2009/14 Y. BARBESEL - P. GIVORD - S. QUANTIN
Partage du chiffre d'affaires, approche par données microéconomiques

G2009/15 I. BUONO - G. LALANNE
The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade

G2010/01 C. MINODIER
Avantages comparés des séries des premières années publiées et des séries des valeurs révisées - Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France

G2010/02 V. ALBOU - L. DAVEZIES - T. DEBRAND
Health Expenditure Models: a Comparison of Five Specifications using Panel Data

G2010/03 C. AFSA - P. GIVORD
La modélisation des comportements démographiques dans le modèle de microsimulation Dynafam

G2010/04 M. BARLET - D. BLANCHET - L. CRUSSON
Social Security and Well-Being of the Elderly: the Case of France

G2010/05 H. RAOUI - S. ROUX
Social Security and Well-Being of the Elderly: the Case of France

G2010/06 R. AEBERTHARDT - J. POUGET
Une approche empirique sur données individuelles

G2010/07 S. BLASCO - P. GIVORD
L’estimation d’un coût implicite de la pénibilité du travail chez les travailleurs âgés

G2010/08 S. QUANTIN - S. RASPILLER - S. SERRAVALLE
Commerce intra-groupe, fiscalité et prix de transferts : une analyse sur données françaises et étrangères

G2010/09 M. CLERC - V. MARCUS
Élasticités-prix des consommations énergétiques des ménages

G2010/10 G. LALANNE - E. POULIQUEN - O. SIMON
Prix de pétrole et croissance potentielle à long terme

G2010/11 B. BELLAMY - G. GONZALEZ - M. FESSEAU - S. LE LAIDER - É. RAYNAUD
Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003

G2010/12 J. BARDAJ - F. TALLET
Detecting Economic Regimes in France: a Quantitative Markov-Switching Indicator Using Mixed Frequency Data

G2010/13 R. AEBERTHARDT - D. FOGUIÈRE - R. RATHLEOT
Démocritus Discrimination à l’emploi : comment exploiter les procédures de testing ?

G2010/14 Y. BARBESEL - P. GIVORD - S. QUANTIN
Partage du chiffre d'affaires, approche par données microéconomiques

G2010/15 I. BUONO - G. LALANNE
The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade