Are retrospective rail punctuality indicators useful? Evidence from users perceptions

Thierry Blayac, Maïté Stéphan

Are retrospective rail punctuality indicators useful? Evidence from users perceptions

Thierry Blayac *
CEE–M, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France
Maïté Stéphan †
Univ Rennes, CNRS, CREM – UMR6211, F-35000, France

January 22, 2021

Abstract

This study analyzes the perceptions of individuals on retrospective rail punctuality indicators to determine the most useful indicator according to socio-demographic characteristics, regular trip behavior variables, and railways transportation habits variables. In choice situations, individuals must choose between four punctuality indicators and an out option. Common punctuality indicators have been selected among those proposed by the authority for quality of service in transport, as well as a new punctuality indicator from the financial literature: Delay-at-Risk. Thus, via an online survey and econometric modeling, we show that respondents appreciate the usefulness of punctuality indicators for planning their long-distance rail trips. The usefulness is reinforced by the fact that respondents employ several modes for regular trips and frequent train users. Moreover, they have already experienced missed appointments or connections. The risk attitude and prudence of respondents also play an important role but not totally in the expected direction. Lastly, Delay-at-Risk, although unknown and more complex in its formulation, exhibits some characteristics that are appreciated by users.

JEL classification:

C25 – R40

Keywords: Reliability; User perceptions; Rail transportation; Retrospective punctuality indicators; Information

* Corresponding author: Thierry Blayac, Faculté d'économie, Site de Richter, Avenue Raymond Dugrand, CS 79 606, 34960 Montpellier Cedex 2, France. Email adress: thierry.blayac[at]umontpellier.fr

† Email address: maite.stephan[at]univ-rennes1.fr

1 Introduction

According to the *Global Carbon Project*[1] (*GCP*), global CO_2 emissions increased in 2018. The analyses and projections made in three recent studies conducted by members of the *GCP* (Figueres et al., 2018; Jackson et al., 2018; Le Quéré et al., 2018) allow Robbie Andrew[2] to state the following: "The slowdown in emissions growth from 2014 to 2016 was always a delicate balance, and the 1.6% increase in 2017 and growth in excess of 2% in 2018, clearly demonstrates that more needs to be done to reduce emissions." Certainly, not all economic activities contribute to equal proportions of CO_2 emissions. Globally, CO_2 emissions from the transport sector represent about 28% of total CO_2 emissions in 2017 and have long been in a steady state. In Europe, the situation is special, since the transport sector accounts for a larger share of total CO_2 emissions (37.5% in France and 40% in Sweden for 2017).[3]

Reducing CO_2 emissions from the transport sector should, therefore, be a priority objective for policymakers. Modal shift towards public transportation or cleaner modes[4] has often been considered the main way to achieve significant reductions in CO_2 emissions from the transport sector (passenger or freight transport) in urban or interurban areas. The notion of modal shift is renewed with the advent of the Flygskam phenomenon[5] in Sweden at the end of 2018 and its spread to the rest of Europe.[6] Nevertheless, the price per kilometer in France is not always in favor of the railway transportation sector. When it is the case, however, it is not always sufficient to generate a lasting change in user behavior. At least, quality is undoubtedly as important as the price signal to induce a change in user behavior and encourage modal shift. Over the past decade, service quality in the railway transportation sector has become a growing issue in Europe for all actors of the transport system (European Directive 1371-2007; ITF, 2017; SNCF, 2012).

The concept of quality is multidimensional, subjective, and highly heterogeneous depending on the transport mode and implicated agents (*e.g.*, passengers and operators). Guirao et al. (2016) employed traditional customer satisfaction surveys to show that, among the many aspects of customer satisfaction, punctuality is ranked in the first position by public transport users. Thus, having information on the punctuality of the various transport modes is an element that may influence modal choice and, therefore, modal shift. Moreover, Grotenhuis et al. (2007) highlight the role of travel information as a key factor of service quality and distinguish three kinds of information: static, dynamic, and real-time. They emphasize the usefulness of these information types for planning and executing trips. Thus, the combination of the two elements (*i.e.*, information and punctuality) serves as a decision tool for the user to plan a trip, with or without connection. The

[1] https://www.globalcarbonproject.org/

[2] Robbie Andrew is a senior researcher at the Center for International Climate Research (CICERO), Oslo, Norway.

[3] The authors compute these figures based on the data provided in the *Fossil CO_2 emissions of all world countries 2018 report* (Muntean et al., 2018).

[4] This study employs the term *modal shift* to refer to the shift from other modes of transport(*i.e., car and air)* to rail.

[5] This term literally refers to *the shame of flying*. This new phenomenon employs modes of transport other than flying to limit greenhouse gas emissions.

[6] For instance, in France, a legislative draft was introduced in June 2019 to replace domestic flights by trains when train travel time is less than 3 h 30 min. However, it should be noted that this legislative draft has not been adopted by the French National Assembly.

question then arises as to what relevant information on punctuality should be provided to users, which this study primarily addresses.

This study specifically focuses on rail punctuality indicators. An online survey is conducted on a representative sample of the French population. Individuals must choose among four punctuality indicators and an out option. In the literature on the reliability of travel times, there are many retrospective indicators of punctuality. In addition to indicators defined by public organizations (*e.g.*, the authority for quality of service in transport, AQST)[7], three common punctuality indicators are selected (*i.e.*, percentage of trains on time (PERCENT), average delay (MEAN), and maximum delay (MAX)). Moreover, a new punctuality indicator from the financial literature is also introduced: Delay-at-Risk (DaR). We then analyze individual perceptions on the indicators to determine the most useful according to socio-demographics characteristics, regular trip behavior variables, and railways transportation habits variables. We estimate a nested dichotomies model that highlights the usefulness of some of the proposed punctuality indicators. Moreover, the econometric modeling allows us to examine the sensitivity of this choice to the various variables introduced in the analysis (*e.g.*, attitude toward risk, regular trip behavior, and railway transportation habits).

The rest of the paper is organized as follows. Section 2 provides the background and the study context as well as a literature review on retrospective rail punctuality indicators. Section 3 presents the questionnaire, the data collection process, and five assumptions to be tested regarding the general usefulness of the retrospective rail punctuality indicators according to individual characteristics (*e.g.*, number of modes used, risk-adverse and prudent behavior, and frequency of using a train). Section 4 provides the sample descriptive statistics and results of the econometric modeling by a nested dichotomies approach with four binary logit models. It also provides comments on the results both statistically and from the decision tool perspective. Section 5 concludes.

2 Background and research context

2.1 Study context

The strategic role of information in guiding/influencing individual and collective behaviors is unanimously recognized by academic researchers from various disciplines such as economics, management, and psychology[8], as well as professionals from the transportation sector (*e.g.*, traffic managers, network planners). This study focuses on passenger information, and more specifically rail punctuality indicators. Due to the ease of access to the data needed to assess the various punctuality indicators, the choice of rail transportation is mainly opportunistic. Our study could be applied to other modes of public transportation (*e.g.*, interurban and urban buses, subway) without any difficulties.

Beyond the results of this study, the ultimate goal of our research program would be to quantify the impact of punctuality indicators on users' modal choice. Nevertheless,

[7]The authority for quality of service in transport (AQST: *Autorité de la qualité de service dans les transports* in French) was created in France in 2012. The AQST ensures the improvement of service quality, including the regularity and punctuality in passenger transportation, information quality provided to passengers in normal circumstances, and degraded or disrupted circumstances.

[8]Of course, this list is not exhaustive. The disciplines mentioned are those to which most of our research is related.

achieving this final goal requires first addressing two main issues to first be addressed: (i) What is the most appropriate source of information (*i.e.*, static, dynamic, real-time)? and (ii) What are the most useful punctuality indicators for travelers? Our study deals with these two issues.

Recent digital innovations (*e.g.*, Big data, Internet of Things (IoT)) have made it possible to access passive, reliable, and real-time information. This real-time information has focused the attention of many researchers over the last few years (Brakewood and Watkins, 2019; Mulley et al., 2017; Brakewood et al., 2014), seeming to relegate static information to the background. If we use the taxonomy provided by Grontenhuis et al. (2007), both dynamic and real-time information can be perceived as particularly useful when individuals are executing a trip (wayside and on-board steps). Meanwhile, static information can be perceived as particularly useful when individuals are planning a trip (pre-trip step). Thus, both dynamic and real-time information are likely to change mobility behavior in the short term. Given the focus on changes in such behavior over time, we operate within a medium-term decision-making horizon; thus, static information is relevant. It is based on retrospective data (for instance, 3 to 6 months) and allows for the computation of punctuality indicators,[9] which is why they are termed retrospective indicators.

The issue of the usefulness of retrospective rail punctuality indicators inevitably refers to how passengers perceive them. To this end, the respondents were placed in hypothetical situations of planning a train trip with and without connection, and were provided with four retrospective punctuality indicators. For each of the choice situations, respondents decided which indicators were most useful for planning the trip. Of the four retrospective rail punctuality indicators proposed, three are well known to users because they are already used by transportation quality agencies or by railway operators on their websites. The last indicator, whose use in a transportation context is totally new, is derived from the financial literature. We call it the Delay-at-Risk (DaR).

This study assesses the usefulness of the various retrospective rail punctuality indicators from the users' perceptions. Among the explanatory factors that we believe may play a role in the choice of one punctuality indicator over another, attitude toward risk on travel time (unreliability) or prudence are probably the most important. To determine these two parameters (*i.e.*, reliability attitude and prudence) for each individual surveyed, we deliberately chose to do so for regular trips with which individuals are familiar. Another key factor of the choice could also be related to the frequency of train use. Nevertheless, since the ultimate goal is also to question the potential of this new indicator DaR to generate modal shifts toward more environmentally friendly modes, there is a clear interest in questioning the perceptions about the usefulness of retrospective rail punctuality indicators for individuals who never use or make little use of rail.

[9] It is conceivable that a user wishing to change the mode of transport on a long-term basis from their personal car to, for example, using trains would want to have information on train punctuality that is established over a sufficiently long period of time such that they can have some confidence in the change. In this context, a 3 to 6 months length for the computation of these indicators appears as a good compromise between computation on a monthly basis and on an yearly basis, which is generally available on the official websites of transportation quality agencies.

2.2 Literature review on retrospective rail punctuality indicators

This section presents a brief literature review on rail punctuality indicators derived from reliability measures and other sources and provides a detailed presentation of the four indicators employed in the survey.

Regarding punctuality indicators, which comprise travel time reliability (or variability) measures, Lomax et al. (2003) and Van Lint et al. (2008) established an exhaustive literature review on such measures for road transportation. They distinguish four main types of measures: statistical, buffer, tardy-trip, and probabilistic measures. **Statistical measures** describe reliability as the travel time variability around its average, which is usually represented by the standard deviation. They are the most common indicator and can be estimated using automatically collected data on roads. **Buffer measures** provide the percentage of travel time to expected to be on time in 95% of the cases, which amounts to one delay in a month. **Tardy-trip measures** put forward the worst possible travel time the maximum delay it may cause. **Probabilistic measures** determine the probability that the travel will take place within a reliable threshold. These measures are usually employed in the road transportation context, especially to identify congestion levels.

Using these measures in a rail transportation framework is difficult because there are precise schedules that must be adhered to. Thus, it is necessary to define punctuality indicators that can be applied to rail transportation. Rietveld et al. (2001) provide a list of six punctuality indicators appropriated for public transportation: (i) the probability that a train arrives x minutes late, (ii) the probability of an early departure, (iii) the difference between the expected arrival and the scheduled arrival time, (iv) average minutes of delay, (v) average minutes of delay for delayed trains, and (vi) standard deviation of arrival time. Van Loon et al. (2001) highlight the main limits of each punctuality indicator. The indicators of average delay ((iv) and (v)) include the usual defaults of average calculations. Namely, there is no difference between a situation that all trains are one minute late, and two out of ten are 5 minutes late (Börjesson and Eliasson, 2011). The calculations of delay probability ((i) and (ii)) do not allow us to differentiate between 3 minutes and 30 minutes. Once the delay threshold has been defined, it will be counted in the same way, whatever the delay size. Moreover, an improvement in service quality cannot be identified if efforts are realized to reduce delays, for instance, a situation in which delays are reduced from 30 to 15 minutes (Van Loon et al., 2011). Nevertheless, this list of punctuality indicators (Rietveld et al., 2011), particularly indicator 1, is used by European countries to establish railway quality indicators.

Since the European regulation n°1371/2007, a quality assessment of the railways is required, especially with regard to delays and cancellations. Nevertheless, this European regulation does not provide any indicators to use, they are left to the analyst's discretion. At the European level, ARAFER[10] (2018) and IRG-Rail (2018) have led studies on monitoring service quality and passengers' rights in railway transportation. We focus only on the results of punctuality indicators developed in 24 out of 27 European countries. ARAFER

[10]The ARAFER (*Autorité de Régulation des Activités Ferroviaires et Routières* in French), which became the ART (*Autorité de Régulation des Transports* in French) in October 2019, is an independent public authority. It was created in 2009 under the name *Autorité de Régulation des Activités Ferroviaires* (ARAF) to accompany the opening of the rail transportation market to competition. Its missions were extended in 2015 and 2016 to include the Channel Tunnel, intercity bus transport, and freeways under concession, making the Authority a multimodal transport regulator.

(2018) highlights differences in the measurement point of delays and the delay thresholds retained to calculate the probability that a train arrives x minutes late. There are three main measurement points, at the final station of service (12 countries), at departure (3 countries), and between each station (8 countries). This last measurement is the better way to determine punctuality trains, but it is requires more calculation. All countries use an indicator that measures the percentage probability that a train will arrive with a delay below a certain threshold (Rietveld et al., 2001). The difference results from the threshold chosen by each country, and also, the thresholds vary in whether they consider the regional or long-distance services. Table 1 presents the delay thresholds for regional and long-distance services in Europe according to Grechi and Maggi (2018) and ARAFER (2018).

Table 1: Delay thresholds for regional and long distance services in Europe

Delay threshold	Regional service	Long-distance service
More than 30 seconds	Hungary	Hungary
More than 1 minute	Croatia	Croatia
More than 2 minutes and 30 seconds	Finland	
More than 2 minutes and 59 seconds	Denmark Switzerland	Denmark Switzerland
More than 3 minutes	Spain[1] Netherlands	Spain[1]
More than 3 minutes and 30 seconds	Latvia	Latvia
More than 3 minutes and 59 seconds	Norway	
More than 5 minutes	Bulgaria United-Kingdom Poland Portugal Slovakia	Bulgaria Netherlands Poland Portugal Slovakia
More than 5 minutes and 29 seconds	Austria	Austria
More than 5 minutes and 59 seconds	Germany Belgium France Sweden	Germany Norway France[1] Sweden
More than 10 minutes		United-Kingdom
More than 10 minutes and 59 seconds		Belgium
More than 15 minutes		Italy[1]
More than 30 minutes	Lithuania	Lithuania

Source: Grechi and Maggi, 2018; ARAFER, 2018
[1] For Spain, France, and Italy, the delay thresholds vary to the category of the journey or travel time.

In France, the AQST developed practical punctuality measures for the railway transportation sector. It defines many punctuality indicators according to arrival or departure delays. We only considered the such indicators on arrival because a departure delay can be recovered and, thus, bears no impact for users. Moreover, such delay will always be present on arrival if it is not recovered. We retain the three main indicators, for a given connection or a set of connections, calculated monthly or average over x months are considered as follows. The first is **the percentage of late trains on arrival**. The second is **the average delay of late trains on arrival** and, finally, **the average delay of trains on arrival**. The difference between the second and the last indicator lies in the fact that the latter includes all trains, whereas the second indicator considers late trains.

We also consider the information provided by SNCF, the French national railway operator. The operator talks about regularity instead of punctuality. However, regularity considers the notion of delay. Regularity is variously defined according to the type of service (*i.e.*, Express Regional, Intercity, and high-speed rail (HSR)). For instance, regarding HSR, the definition of regularity depends on the travel time of passengers. A train is considered to be on time if the delay is less than 5 (15) mins for a travel time that is less (greater) than 1h30 mins (3 h). From these thresholds, SNCF determines a percentage of train regularity that it is equivalent to a percentage of train punctuality.

Finally, we considered the financial literature on risk measures, assuming that the lack of punctuality impacts travel time. That is, a delay can be viewed as a potential loss for a journey, which can be explained by financial measures of risk such as value-at-risk (VaR). VaR is defined as the loss level that will not be exceeded with a specified probability (Hull, 2015). Therefore, we propose a punctuality indicator inspired by VaR that adapts to rail transport: DaR. DaR is the upper limit of the delay that should only be exceeded within a given probability.[11] Thus, the number of rail punctuality indicators is limited to the four indicators in Table 2. These indicators are grouped into two sub-groups: the central tendency indicators with PERCENT and MEAN indicators, and the extreme risk indicators with MAX and DaR indicators. We then investigate the usefulness of the information provided by the indicators of rail punctuality to individuals in planning their trips. Hence, we conducted a survey of a representative sample of the French population.

[11]For a formal definition, see the Appendix. For more details regarding definitions and estimations, see Mbairadjim Moussa and Stéphan (2014).

Table 2: Retrospective rail punctuality indicators employed in the survey

	Notation	Definition	Example
Central Tendency	PERCENT	PERCENT gives the rate of trains on time on arrival.	From Montpellier to Paris, 74% of trains are on time.
	MEAN	MEAN gives the average delay of all trains and the average delay of late trains.	From Montpellier to Paris, the average delay is 5 minutes for all trains, and is 27 minutes for all late trains.
Extreme risk	MAX	MAX gives the maximum delay in minutes encountered during a given period and a given link. It is part of the tardy-trip measure.	From Montpellier to Paris, the maximum delay is 192 minutes (3 h 12 mins).
	DaR	Delay-at-Risk (DaR) gives the upper limit of the delay that should only be exceeded with a given probability.	From Montpellier to Paris, a train will have a delay greater than 23 minutes in 5% of cases.

3 Methodology

An empirical strategy based on a questionnaire survey is implemented to answer the research question. Section 3.1 discusses the structure of the questionnaire, the data collection method, and the definition of some important variables. Section 3.2 establishes assumptions to be tested using the data collected on our representative sample of the French population.

3.1 Survey and data collection

3.1.1 Questionnaire structure

The questionnaire framework is organized into three parts and comprises 27 items. The first part lists traditional socioeconomics and demographics items, such as gender, age, household income, and household structure (*e.g.*, child, number of individuals). The second part addresses regular trip behavior (*e.g.*, schedule, transport mode, travel time, and cost). The questionnaire is unusual since some questions are personalized regarding travel time experienced by users during a regular trip.[12] Thus, personalized lotteries on travel time are deduced to determine the attitude toward travel time reliability for each individual (Beaud et al., 2016). The last part of the questionnaire investigates perceptions related to different punctuality indicators. Respondents are subjected to four framing contexts and must choose, in each context, the indicator they feel is the most useful. Thus, to avoid any anchoring effects, the indicators appear randomly, with an out option is available. This part ends on the subject of the habit of respondents in using train

[12] A regular trip is a daily trip or almost daily trip, such as commuting trips, home-to-study trips, shopping trips, or leisure trips. We did not want to limit the context to commuting trips because the risk is to exclude unemployed and retired people.

transportation (*e.g.*, frequency of using a train and missed appointment and connection). Figure 1 provides a screenshot of a choice situation.

You must book a ticket for an unconnected train. The railway operator announces a travel time of 3 h 20 mins. The booking website gives the following indicators:

- ☐ Indicator 1: The rate of trains that are on time is 74%.
- ☐ Indicator 2: The average delay is 5 mins for all trains, and is 27 mins for all late trains.
- ☐ Indicator 3: In 5% of cases, the train will have a delay higher than 23 mins.
- ☐ Indicator 4: The maximum delay is 192 mins.
- ☐ Indicator 5: No information is useful.

Please, choose the best indicator for you.

Figure 1: Screenshot of a choice situation

The punctuality indicators employed in the survey are presented in Table 2. In Figure 1, the first indicator corresponds to the percentage of trains on time (PERCENT). The second indicator is the average delay (MEAN). The third indicator is DaR. The fourth indicator is the maximum delay (MAX). Finally, the fifth indicator is the out option. The respondents are subjected to four choice situations for which both punctuality indicators and durations are always identical.[13] The difference between the choice situations lies in the connection time (*i.e.*, 0, 5, 30 et 60 mins). For instance, in Figure 1, the connection time is 0 mins. [14].

3.1.2 Data collection method

The sample was built by employing the services of Survey Sampling International Inc. (SSI)[15] to guarantee its representativeness. Regarding the sampling method, we employ the quota method based on the following criteria: gender, age, and income. Data collection was implemented via an online survey using the LimeSurvey software from February 15 to 23, 2018. We obtained a total of 670 completed surveys. Moreover, 536

[13] These indicators have been calculated via data collected from July to September, 2012 between Montpellier to Paris using an HSR service that travels both ways. For more details, see Mbaraidjim Moussa and Stephan, 2014. They express and translate, each in their own way, the same level of punctuality of the HSR link studied.

[14] An English translation of the French questionnaire used for the survey is available in the Appendix 6.2. More specifically, the four various choice situations proposed to individuals are presented as items 13 to 16.

[15] All information about SSI procedures and the pool of respondents may be found at the following URL address: `www.surveysampling.com`. SSI has an international pool of respondents that are selected and targeted depending on the requirements of researchers. In our case, the online questionnaire was published on the French platform and members of the French pool could connect and fill the survey if they were preselected by SSI's algorithm calibrated to implement our quotas. Since our study began, SSI Inc. has become part of Dynata Corporation.

incomplete surveys did not fulfill our requirements, and, thus, were excluded from the final sample by SSI. Despite the specificity of the questionnaire (see, for instance, items 12 to 16 in the Appendix), respondents did not experience too many cognitive difficulties. Of our 670 respondents, only 44 (6.57%) declared having some difficulties to answer the survey.[16] In other words, 93% of respondents had no difficulties. The individuals with difficulties in answering are largely those who have not or rarely an experience of delay. Table 3 provides the sample characteristics.

Table 3: Sample descriptive characteristics

Variable	Percentage (%)	Variable	Percentage (%)
Age		**Family situation**	
18 to 24	11.79 **[11]**	Single	23.13
25 to 34	15.67 **[16]**	In a couple	67.76
35 to 44	19.85 **[18]**	Others (divorced, widowed)	9.10
45 to 54	19.10 **[18]**		
55 to 64	15.22 **[15]**	**Children**	
Above 64	18.36 **[22]**	Yes	61.64
		No	38.36
Gender			
Male	46.72 **[48]**	**Transportation mode**	
Female	53.28 **[52]**	Car and motorbike	68.51 *[72]*
		Train	7.61 *[4]*
Income		Bus	7.31 *[4]*
Below €1000	9.25 **[10]**	Tramway	4.18 *[2]*
€1001 to €1500	9.10 **[10]**	Biking	1.94 *[6]*
€1501 to €2500	30.00 **[30]**	Walk	8.51 *[1]*
€2501 to €3500	10.15 **[10]**	Subway	1.64 *[9]*
€3501 to €5000	20.75 **[20]**	Others	0.3
€5001 to €6500	10.30 **[10]**		
Above €6500	10.45 **[10]**		

Notes:

- The quota's criteria targeted (*i.e.*, age, gender, and income) are specified in brackets and in bold.
- For the transportation mode variable, the numbers in brackets and in italics correspond to the results of the study conducted by BVA (2015).

The χ^2 tests of adjustment[17] confirm that the sample is representative of the French population regarding age, gender, and income distribution. Concerning the transportation mode, the sample is not representative even if the gap between the observed and real proportion is low. For instance, 69% of respondents used car or motorbike for their usual trips, as compared to 72% of the French population, according to BVA (2015). The use of transportation modes was not a targeted criterion for the survey. However, regarding the gap between our results and BVA (2015), the latter allows individuals to choose many transportation modes, whereas respondents must choose an exclusive transportation mode in this study.

[16] Among the difficulties mentioned, some are related to the unusual formulation of certain (personalized) questions, difficulties in projecting oneself into a regular trip given one's situation or the non-use of the train, or difficulties related to understanding certain statistical concepts.

[17] See supplementary materials - Section 1

3.1.3 Variables definition

Based on the implemented survey, we define 16 qualitative variables. To simplify the explanation of variables, we group them under three categories: socio-demographics, regular trip behavior, and railways transportation habits variables.

First, socio-demographic variables are defined as gender, age, income, family situation (*i.e.*, single, couple, and other), and the presence of a child. For the age variable, four categories were created: the first includes many students (*i.e.*, 18–24 years), the second is mainly made up of young workers (*i.e.*, 25–34 years), the third mainly comprises experienced workers (*i.e.*, 35–64 years), and the last corresponds to older people (*i.e.*, 65 years and more).

Seven variables concerning regular trip behavior are also defined. In the survey, each respondent must describe their regular journey[18] (*e.g.*, commuting, leisure journey). Many of these variables are common, such as the travel time (TT), the perception of travel cost (COST), the main mode used (MODE), and the number of modes used (NMB). NMB allows for identifying individuals employing multimodal transportation for their regular trips. We also use the delay frequency (DELAY) to capture respondents' experience of delay. It defines whether an individual is in the habit of being late due to transport disruptions.

Even so, based on information from the survey, it is possible to compute a safety margin (SM) for each individual. Knight (1974) was the first to introduce the safety margin concept. He defines it as the additional travel time an individual integrates with the total travel time to cope with the negative consequences of an uncertain travel time. More recently, Tam et al. (2008) explain the safety margin as the difference between preferred and expected arrival time. Börjesson et al. (2012) describe it as the difference between time constraints at the destination and scheduled arrival time. We adopt Börjesson et al.'s (2012) definition as follows:

$$\text{SM} = t_w - t_h - E(t), \tag{1}$$

where t_w is the start time of activity, t_h is the departure time from home, and $E(t)$ is the expected (or average) travel time. From Equation 1, three cases can occur. First, $\text{SM} < 0$ implies that an individual does not plan to leave early enough to arrive on time regarding predicted travel time. Second, $\text{SM} = 0$ implies that an individual does not insure against a possible delay. Third, $\text{SM} > 0$ implies that an individual plans a safety margin to ward against an uncertain travel time.

Finally, the last variable in the second category is the individual reliability attitude (RELIABATT). We use the definition provided by Beaud et al. (2016): "a traveler is reliability-prone whenever he/she always prefers a reliable trip with a certain travel time to any risky trip with a random travel time, whenever both trips feature the same cost and the same mean travel time.". Thus, to determine the individual reliability attitude, each respondent is subjected to a lottery, as represented by Figure 2.
The average travel time declared by a respondent is $\bar{t}$. Using $\bar{t}$ allows for proposing a lottery on travel time close to an individual's habit because the proposed lottery depends directly on individual usual travel time ($\bar{t}$). A lottery with a certain travel time is defined

[18]There are not restrictions regarding the individuals who can describe their journeys. For instance, unemployed persons and retired persons are not excluded.

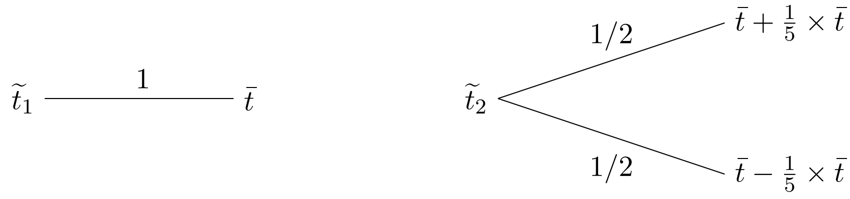

Figure 2: Lottery on travel time to determine reliability attitude

by $\tilde{t}_1$, whereas $\tilde{t}_2$ is a lottery with a random travel time. In the survey, the random travel time is defined for each respondent as his average travel time plus or minus 20% with a 0.5 probability. According to the choice between $\tilde{t}_1$ and $\tilde{t}_2$, the individual reliability attitude can be determined as follows. (i) If an individual prefers $\tilde{t}_2$ to $\tilde{t}_1$ ($\tilde{t}_2 \succ \tilde{t}_1$), the individual is reliability-averse. That is, the individual prefers a journey with a random travel time over a journey with a certain travel time. (ii) If an individual prefers $\tilde{t}_1$ to $\tilde{t}_2$ ($\tilde{t}_1 \succ \tilde{t}_2$), the individual is reliability-prone. It means that an individual prefers a journey with a certain travel time to a journey with a random travel time. (iii) If an individual is indifferent between $\tilde{t}_1$ and $\tilde{t}_2$ ($\tilde{t}_1 \sim \tilde{t}_2$), the individual is reliability-neutral. The individual is indifferent between a journey with a certain travel time and one with a random travel time (with the same average travel time, of course).

Finally, the last category of variables includes all variables corresponding to train trip habits for individuals: the frequency of train transport employed (FREQ) and whether individuals have already missed a connection (CONNECT) or an appointment (APT) because of train delay. The last variable defined is the connection time (CONNTIME). It takes the value of 0, 5, 30, or 60 mins according to the submitted questionnaire. We recall that there are four questionnaires about the choice of punctuality indicators. Table 4 presents all qualitative variables defined in the study with their various categories and their corresponding notations.

3.2 Assumptions to be tested

The issue of the usefulness of retrospective punctuality indicators to plan long-distance railway trips leads us to propose several hypotheses. This section presents five hypotheses to test using the data collected in our survey.

- **Hypothesis 1**. Users find retrospective punctuality indicators to be useful for planning their long-distance railway trips.

Among all the proposed indicators, the first question to be answered is whether they are useful for users. This hypothesis can be tested because an out option is provided in the survey for the four choice situations. To the extent that this information is freely available, individuals may be more inclined to use it to plan their long-distance railway trips. Furthermore, it might be interesting to know whether the useful perspective of punctuality indicators is associated with some socio-economic characteristics, with regular trips behavior, train use habits.

Table 4: Variable definition: Notation and categories

Socio-demographics		Regular trip behavior		Railways transportation habits	
Variables	Categories	Variables	Categories	Variables	Categories
Age	18–24 years	**Mode**	Motorized vehicles	**Frequency**	Never
(AGE)	25–34 years	(MODE)	Collective transport	**of use**	Yearly
	35–64 years		Soft mode	(FREQ)	Quarterly
	65 years and more				Monthly
		Number	NBM1= 1		Weekly
		of modes	NBM2 = 2 or 3		Daily
		(NBM)	NBM3 = 4 or 5 or 6		
Gender	Male			**Connection**	Yes
(GEN)	Female	**Travel**	TT1 $\in [1;15]$	**missed**	No
		time	TT2 $\in [16;30]$	(CONNECT)	
Family	Single	(TT)	TT3 $\in [31;45]$		
situation	Couple		TT4 $\in [46$ and $+[$	**Appointment**	Yes
(FAM)	Others			**missed**	No
		Cost	COST1 $\in [0, 1.25]$	(APT)	
Income	€0-1000	(COST)	COST2 $\in]1.25, 3]$		
(INC)	€1001-1500		COST3 $\in]3, 5]$		
	€1501-2500		COST4 $\in]5$ and$+[$	**Connection time**	0 min
	€2501-3500			(CONNTIME)	5 mins
	€3501-5000	**Safety**	SM1 $\in [-30; 0[$		30 mins
	€5001-6500	**margin**	SM2 $\in [0]$		60 mins
	Above €6501	(SM)	SM3 $\in [1; 10[$		
			SM4 $\in [10; 15]$		
Child	Yes		SM5 $\in [16$ and $+[$		
(CHILD)	No				
		Reliability	Prone		
		attitude	Neutral		
		(RELIABATT)	Averse		
		Delay	Never		
		frequency	Rarely		
		(DELAY)	Sometimes		
			Frequently		
			Very frequently		

- **Hypothesis 2**. Users using several modes in their regular mobility are more likely to appreciate having a retrospective punctuality indicator to plan their long-distance railway trips.

Behind this assumption is the fact that individuals who use several modes of transport for their daily mobility needs are more inclined to use information sources or punctuality indicators to avoid load breaks or disruptions. They would not change their habits when they have to plan a long-distance railway trips.

- **Hypothesis 3**. Risk-averse and prudent individuals should be more inclined to use rail punctuality indicators to plan their long-distance railway trips.

In the economics literature, many studies (Von Neumann and Morgenstern, 1944; Pratt, 1964; Kahneman and Tversky, 1979; Kimball, 1990; Tversky and Kahneman, 1992)[19] have

[19]We mention only the most salient examples and without being exhaustive.

highlighted the role played by risk attitudes and prudence in the decision-making process of economic agents in a risky environment. From a transportation context, the reliability proneness (Beaud et al., 2016) is the risk-averse attitude measure, while the safety margin (Knight, 1974) is a proxy of the prudence concept. That is, a reliability-prone individual dislikes the risk on the travel time and employ the safety margin ward against the risk. To avoid any negative issues due to an unreliable transport mode, a risk-averse and prudent user will probably use retrospective punctuality indicators to plan long-distance railway trips.

- **Hypothesis 4**. Frequent train users are more likely to appreciate having a rail punctuality indicator than occasional users.

This hypothesis seems quite logical and could be described as the "strength of habits." In France, the national railway transport operator (SNCF) does not have a great reputation for reliability, unlike, for instance, the Japanese operators. This reliability varies according to the routes considered. Thus, frequent train users have arguably become accustomed to using information sources or retrospective punctuality indicators before planning any long-distance railway trips.

- **Hypothesis 5**. Missing an appointment or a connection should encourage users to choose a rail punctuality indicator and, more specifically, extreme risk indicators (MAX and DaR).

Arguably, individuals who have already experienced negative events due to train unreliability will be more inclined to use one of the proposed punctuality indicators in choice situations to avoid a repeat of the bad experience. We can legitimately think that these individuals would then prefer indicators that reflect extreme risks than average risks. The phenomenon of pessimism bias supports this hypothesis; it is well documented in the economic and cognitive psychology literature (Ben Mansour et al., 2006; Blanton et al., 2001; Chang and Asakawa, 2003).

4 Data analysis

This section analyzes the data collected through the online survey to generate insights into the usefulness of retrospective rail punctuality indicators for users. More specifically, the various assumptions formulated in the section 3.2 are tested. The methodology involves both descriptive statistics and econometric modeling.

4.1 Overall descriptive results

The first part of the results comprises descriptive statistics on the raw data from the survey. When necessary, these initial statistical analyses are supplemented by tests of association (chi-square) to provide in-depth explanations. As all the variables are qualitative, we employ frequency and percentage statistics to describe them. These statistics are only computed based on complete questionnaires that correspond to the target population (*i.e.*, 670 individuals). Nevertheless, regarding the punctuality indicator choice situations, we aggregate the results for the four proposed situations (*i.e.*, 0, 5, 30, and 60

Table 5: Overall descriptive results

Socio-demographics				Regular trip behavior				Railways transportation habits			
Variables		Obs.	Freq.	Variables		Obs.	Freq.	Variables		Obs.	Freq.
Age	18–24 years	79	11.79	**Mode**	Motorized vehicles	461	68.81	**Frequency**	Never	146	21.79
(AGE)	25–34 years	105	15.67	(MODE)	Collective transport	139	20.75	**of use**	Yearly	176	26.27
	35– 64 years	363	54.18		Soft mode	70	10.45	(FREQ)	Quarterly	123	18.36
	65 years and more	123	18.36						Monthly	123	18.36
				Number	NBM1= 1	452	67.46		Weekly	36	5.37
				of modes	NBM2 = 2 or 3	191	28.51		Daily	66	9.85
				(NBM)	NBM3 = 4 or 5 or 6	27	4.03				
Gender	Male	313	46.72					**Connection**	Yes	232	34.63
(GEN)	Female	357	53.28	**Travel**	TT1 $\in [1; 15]$	202	30.15	**missed**	No	438	65.37
				time	TT2 $\in [16; 30]$	250	37.31	(CONNECT)			
Family	Single	155	23.13	(TT)	TT3 $\in [31; 45]$	90	13.43				
situation	Couple	454	67.76		TT4 $\in [46$ and $+[$	128	19.10	**Appointment**	Yes	293	43.73
(FAM)	Others	61	9.10					**missed**	No	377	56.27
				Cost	COST1 $\in [0, 1.25]$	168	25.07	(APT)			
Income	€0-1000	62	9.25	(COST)	COST2 $\in]1.25, 3]$	218	32.54				
(INC)	€1001-1500	61	9.10		COST3 $\in]3, 5]$	125	18.66				
	€1501-2500	201	30.00		COST4 $\in]5$ and$+[$	159	23.73	**Indicator**	NONE	876	32.69
	€2501-3500	68	10.15					**Choice**	PERCENT	509	19.00
	€3501-5000	139	20.75	**Safety**	SM1 $\in [-30; 0[$	27	4.03	(CHOICE)	MEAN	691	25.78
	€5001-6500	69	10.30	**margin**	SM2 $\in [0]$	208	31.04		MAX	148	5.52
	Above €6501	70	10.45	(SM)	SM3 $\in [1; 10[$	97	14.48		DaR	456	17.01
					SM4 $\in [10; 15]$	176	26.27				
Child	Yes	413	61.64		SM5 $\in [16$ and $+[$	162	24.18				
(CHILD)	No	257	38.36								
				Reliability	Prone	312	46.57				
				attitude	Neutral	181	27.01				
				(RELIABATT)	Averse	177	26.42				
				Delay	Never	252	37.61				
				frequency	Rarely	197	29.40				
				(DELAY)	Sometimes	117	17.46				
					Frequently	80	11.94				
					Very frequently	24	3.58				

Notes: The variable CONNTIME has not been inserted in the Table because it is a choice situation variable; therefore, its number of occurrences is identical for each category (*i.e.*, 0, 5, 30 and 60 mins) and corresponds to the number of respondents (*i.e.*, 670 individuals).

mins of connection time). Thus, the sample size is 2,680 observations. Table 5 presents the results and allows us to discuss the hypothesis developed in section 3.2.

First, we observe that when individuals are confronted with the four punctuality indicator choice situations, in more than two-thirds of choice situations, a punctuality indicator is chosen. Nevertheless, with a third of the responses of choice situations, the out option (NONE) is by far the most popular answer. This high proportion gives the first important result: the provided information (in particular, the information provided by common indicators) does not convince respondents. Moreover, only 5% of choices concern MAX. It, thus, indicates that it is not appropriate to provide punctuality information in standard real-life situations. The indicators MEAN, PERCENT, and DaR were chosen respectively in 26%, 19%, and 17% of cases, respectively. The high frequencies of MEAN and PERCENT indicators were not surprising since they are common indicators. DaR, was chosen in 17% of the choice situations, which is very satisfactory, as compared to the common indicators. Despite its more complex formulation than the other indicators, the DaR demonstrated a good receptiveness from the respondents.

If we assume that an indicator was chosen because of its usefulness in planning long-distance rail trips, then the previous elements seem to validate ***Hypothesis 1***. We can investigate further to establish associations between the fact that a punctuality indicator was found to be useful for planning a long-distance railway trip and socio-demographic characteristics, regular trip behavior, and railway transportation habits. Using Chi-square tests (see supplementary materials), we detected associations between users' perception of the usefulness of punctuality indicators via variables such as family situation, income, travel time, and travel cost. The econometric results provided in section 4.2 allow for inferring or confirming the initial results by refining the direction of the effects according to the various indicators.

Most respondents use only one mode of transportation (67%) for their regular trips. Moreover, this single mode of transportation is the car for 79% of them.[20] Even so, 33% of respondents use multimodal transportation for their regular trips (Table 5). Of the 33% of choice situations where no indicator is considered as relevant by the users, 76% are explained by individuals using only one mode of transportation.[21] Moreover, respondents using only one mode of transportation choose a punctuality indicator in 63% of the choice situations, whereas this proportion increases to 75% for individuals who use two or three modes of transportation and even 85% for those using more than three modes of transportation for their regular trips.[22] Furthermore, respondents using a collective mode of transportation choose punctuality indicators more frequently than respondents using a private mode of transportation.[23] The results are consistent with ***Hypothesis 2***.

Reliability-neutral respondents comprise the lowest proportion of those who find the proposed punctuality indicators useful for planning long-distance railway trips (61%). This proportion increases to 68% for reliability-prone individuals and 73% for reliability-averse individuals.[24] The attitude towards risk regarding travel time, therefore, has an impact on whether individuals find a proposed punctuality indicators to be useful in choice situations. However, the impact is even more pronounced for reliability-averse individuals. If we consider respondents with a strictly positive safety margin as prudent from an economic theory perspective, then the influence of this characteristic on finding retrospective indicators of punctuality useful for planning long-distance railway trips is noteworthy to examine. Nevertheless, we do not observe a major difference. About 65% of individuals choose a punctuality indicator regardless of the value of their safety margin.[25] In any case, the Chi-square test does not reveal any association between prudence and the usefulness of retrospective punctuality indicators for users.[26] The preliminary results seem negative ***Hypothesis 3***, which will have to be confirmed via the econometric analysis in section 4.2.

Hypothesis 4, which evokes the strength of habits, where frequent train users are more likely to appreciate having a retrospective punctuality indicator than occasional users for planning their long-distance train trips, is confirmed. The more trains are used regularly, the more retrospective punctuality indicators are chosen in the choice situations (*i.e.*, the

[20] See supplementary materials - Table 2.
[21] See supplementary materials - Table 3.
[22] See supplementary materials - Table 3.
[23] See supplementary materials - Table 4.
[24] See supplementary materials - Table 5.
[25] See supplementary materials - Table 6.
[26] See supplementary materials - Table 1.

less the out option is chosen by individuals). Only 54% of respondents who never use the train for their regular trips find one of the punctuality indicators useful, as compared to 69% of those who use it quarterly and 81% of those who use it daily.[27] The usefulness of punctuality indicators for the respondents increases with the frequency of using a train. These figures are corroborated by the Chi-square test that allows for the conclusion that there is an association between the usefulness of indicators and the frequency of using a train.[28]

We deliberately included two items[29] in the questionnaire to report any experiences that passengers may have had due to train delays. These negative experiences are related to a missed connections or appointments. The economic and psychological literatures highlights the existence of a pessimistic bias,[30] which can have a great influence on current and future decision-making processes of individuals. In the four choice situations, respondents who have never missed a connection or an appointment due to a train delay choose the out option in the proportions of 38% and 39%, respectively. Meanwhile, individuals who have already encountered such negative events choose the out option in the proportions of 23% and 25%, respectively.[31] Consequently, respondents (76%) who have encountered such negative situations choose a punctuality indicator more often, as compared to those (61%) who have not encountered them. This is a potential illustration of the pessimism bias. The results are confirmed by the Chi-square association test, which supports ***Hypothesis 5***. Nevertheless, descriptive statistics alone is not sufficient to refine the analysis of whether individuals choose punctuality indicators regarding extreme risk (MAX and DaR).

The results obtained must be confirmed in-depth by providing an indicator-by-indicator analysis, which the following the econometric analysis achieves.

4.2 Econometric analysis

This section first presents general considerations on the econometric strategy used in the study. The econometric results are then provided and discussed from a statistical point of view, from which we confirm the various assumptions established, indicator by indicator. Finally, it leads to recommendations and policy implications.

4.2.1 Overview of the econometric strategy

We model the probability that an individual, planning a long-distance trip by train, with or without connection, will find one of the four punctuality indicators useful. In this choice experience, individuals also had an out option; it means that none of the indicators seemed useful to them. More precisely, it is equivalent to econometrically modeling the probability of choosing between five alternatives (*i.e.*, PERCENT, MEAN, MAX, DaR and NONE).

[27] See supplementary materials - Table 7.
[28] See supplementary materials - Table 1.
[29] See items 18 and 19 in Appendix 6.2
[30] Among the elements often put forward to explain this bias of pessimism in the behavior of individuals, having experienced negative events is a possible explanation.
[31] See supplementary materials - Table 8 and Table 9.

We first employed a generalized logit model to estimate this probability, but the Small and Hsiao test (1985) rejected the null hypothesis of the independence of irrelevant alternatives. We then turned to a multinomial probit model, which implicitly assumes the homogeneity of slopes for each of the utility functions associated with the alternatives considered. Since the slope homogeneity test was rejected, we finally opted for the method of nested dichotomies (see Fox, 2016; Friendly and Meyer, 2015). This method consists of separating the m-alternatives of the polytomous variable into m-1 dichotomies, which will each be estimated by a binary logit model. As Fox (2015) points out, this method is reasonable if the choice of dichotomies is not arbitrary. The following nested set of four dichotomies appears to be well-founded. At the first level, we distinguish between choice situations for which a punctuality indicator has either been chosen or not. At the second level, only for situations in which individuals have chosen a punctuality indicator, we study the choice between a central tendency indicator (PERCENT, MEAN) or an extreme risk indicator (MAX, DaR). At the third level, depending on the branch of the tree under consideration, we study the choice between PERCENT or MEAN indicators (Level 3A) or MAX or DaR indicators (Level 3B). These nested dichotomies are illustrated in the Figure 3.

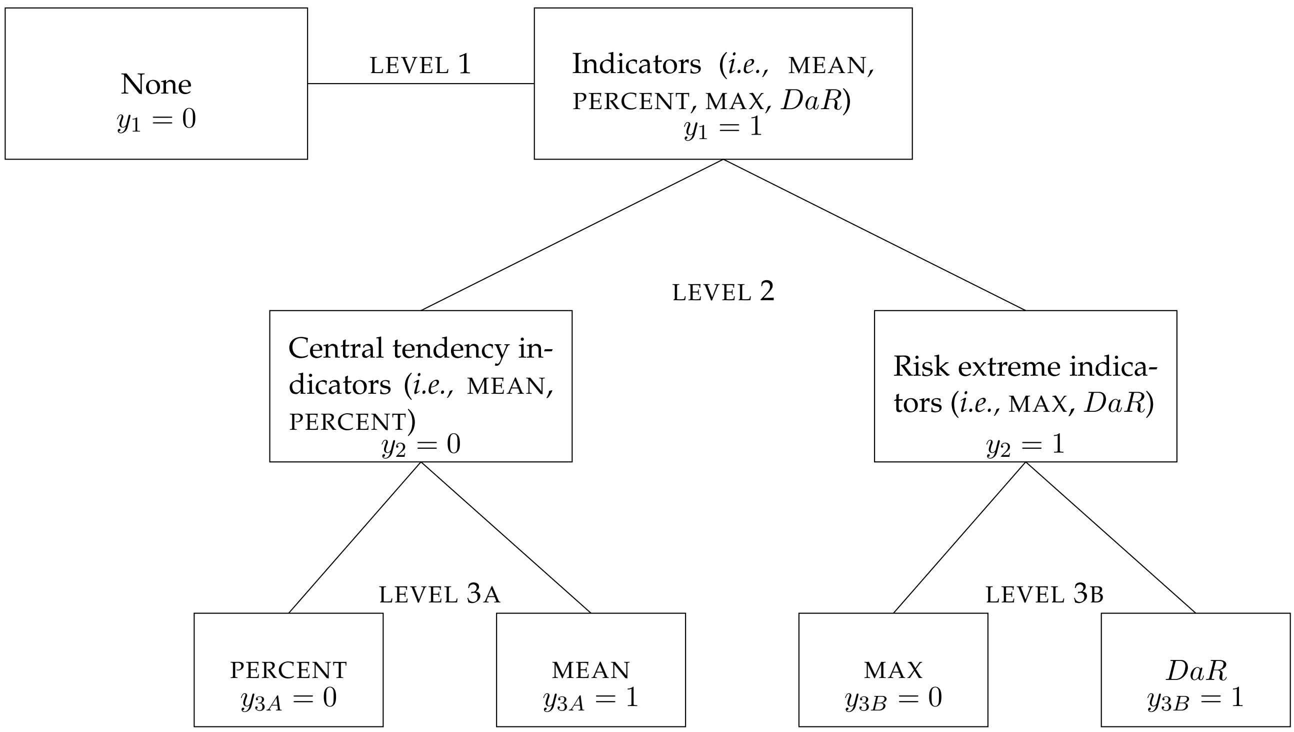

Figure 3: Nested dichotomies structure

At each level, we model the choice between two alternatives with a binary logit. Thus, we estimate four binary logits. As demonstrated in Friendly and Meyer (2015) and Fox (2016), each dichotomy is independent. Consequently, the probability of a given alternative or option is simply the product of the probability obtained at each level, provided that the given alternative exists at the level under consideration. For instance, the probability associated with the feeling that DaR is a useful punctuality indicator to plan long-distance trips by train is given by the following expression:

$$\mathrm{Prob}(DaR) = \mathrm{Prob}(y_{3B} = 1) \times \mathrm{Prob}(y_2 = 1) \times \mathrm{Prob}(y_1 = 1). \quad (2)$$

We employ the estimation strategy described in the Figure 3. Moreover, section 4.2.2 presents the results.

4.2.2 Econometric results

According to our econometric strategy, we estimate four binary logit models regarding the nested dichotomies structure of Figure 3. Table 6 presents the econometric results. At each level of the nested structure, the estimated models are globally valid since the likelihood ratio tests allow for rejecting the null hypothesis (p-value < 0.0001). We can also assess the goodness of fit by considering the proportion predicted with success; this proportion varies from 58.6% to 74.4% depending on the level considered. Thus, considering the entire model, the proportion predicted with success is about 64%.[32] Hence, this result is quite satisfactory.

We now consider the significance of the explanatory variables introduced in the various models. Among the 16 qualitative variables, 13 are statistically significant, at least at the 5% level, according to the type 3 effects analysis.[33] However, they do not operate at the same level of the nested structure. These variables are broadly the same as those whose influences have already been highlighted in the overall analysis of the results (4.1). Section 4.2.3 presents a detailed analysis of the effect of each variable on the choice of a specific punctuality indicator.

4.2.3 Results Analysis

This section tests the assumptions set out in section 3.2. The descriptive statistics have already clarified a few points, but we wish to investigate further by testing the sensitivity of each indicator to the different explanatory variables. Thus, we compute the impact of the different categories of the explanatory variables on the probability of choosing one of the five indicators. Based on the results obtained (see Table 7), we provide some policy implications and recommendations.

Some preliminary comments are useful to understand the results established in Table 7. For the reference category of each explanatory variable, the probability of choosing a given indicator is in bold. These probabilities are computed at the average point of our sample, apart from the reference category. The others figures describe the evolution of these probabilities for any other categories than the reference category, all things being equal. The average point of our sample is characterized for each variable, by the category with the highest frequency (see Table 5). Thus, the average individual is aged 35 to 64 years. He lives in couple with an income range from 1501 to 2500 euros. For his regular trip, he mainly uses his own car; he has no safety margin and is reliability-prone. His perceived travel cost is ranges from 1.25 to 3 euros. He travels once a year by train and, has never missed a connection and/or an appointment due to a train delay.

[32] This figure is obtained from the combination of the proportion predicted with success at each level.

[33] The analysis of type 3 effects makes it possible for a qualitative variable to assess whether the variable exerts an overall effect on the variable to be explained. This may be the case even if some categories of the variable are not significant. The three variables that are not significant are gender, child, and delay frequency.

Table 6: Econometric results

		Level (1)		Level (2)		Level (3A)		Level (3B)	
Dependent Variables		One indicator		Risk Extreme		MEAN		DaR	
Explanatory Variables		coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value
INTERCEPT		0.3642	0.2068	0.3665	0.1391	-0.2691	0.2316	-0.6466	0.1364
AGE	18-24 years	-	-	-0.6101	0.0015	-	-	-0.1485	0.6742
	25-34 years	-	-	ref.	ref.	-	-	ref.	ref.
	35-64 years	-	-	-0.5668	<.0001	-	-	0.5813	0.0263
	65 years and +	-	-	-0.5928	0.0006	-	-	1.1524	0.0039
FAMILY SITUATION	Single	-	-	-	-	-	-	ref.	ref.
	Couple	-	-	-	-	-	-	1.0837	<.0001
	Others	-	-	-	-	-	-	0.8037	0.0889
INCOME	€0-1000	-	-	-	-	ref.	ref.	ref.	ref.
	€1001-1500	-	-	-	-	1.2625	<.0001	0.2247	0.6082
	€1501-2500	-	-	-	-	0.6420	0.0043	0.9660	0.0144
	€2501-3500	-	-	-	-	0.8848	0.0018	0.9081	0.0853
	€3501-5000	-	-	-	-	0.7850	0.0010	-0.2642	0.5249
	€5001-6500	-	-	-	-	0.7243	0.0082	-0.0564	0.9044
	Above €6501	-	-	-	-	0.6551	0.0119	0.1467	0.7769
MODE	Motorized vehicles	-0.2703	0.0810	-0.4801	0.0081	-	-	-	-
	Collective transport	-0.4490	0.0132	-0.2586	0.1953	-	-	-	-
	Soft mode	ref.	ref.	ref.	ref.	-	-	-	-
NBM	NBM1 = 1	ref.	ref.	-	-	-	-	ref.	ref.
	NBM2 = 2 OR 3	0.288	0.0122	-	-	-	-	-0.9072	0.0002
	NBM3 = 4 TO 6	0.9127	0.0019	-	-	-	-	0.3833	0.5704
TT	TT1 $\in$ [1;15]	-0.3667	0.0308	-	-	-	-	-	-
	TT2 $\in$ [16;30]	-0.3038	0.0466	-	-	-	-	-	-
	TT3 $\in$ [31;45]	-0.5965	0.0004	-	-	-	-	-	-
	TT4 $\in$ [46 and +]	ref.	ref.	-	-	-	-	-	-
COST	COST1 $\in$ [0;1.25]	-0.4578	0.0021	-0.3438	0.0337	-	-	-	-
	COST2 $\in$]1.25; 3]	-0.1494	0.2652	0.0475	0.7222	-	-	-	-
	COST3 $\in$]3; 5]	-0.1322	0.3601	0.4185	0.0050	-	-	-	-
	COST4 $\in$]5 and + [	ref.	ref.	ref.	ref.	-	-	-	-
SAFETY MARGIN	SM1 $\in$ [-30; 0[	ref.	ref.	-	-	-	-	-	-
	SM2 $\in$ [0]	0.3879	0.0887	-	-	-	-	-	-
	SM3 $\in$ [1; 10[	0.2877	0.2365	-	-	-	-	-	-
	SM4 $\in$ [10; 15]	0.4977	0.0330	-	-	-	-	-	-
	SM5 $\in$ [16 and +[	0.2184	0.3476	-	-	-	-	-	-
RELIABILITY ATTITUDE	Prone	0.4250	0.0004	-	-	-	-	-	-
	Neutral	ref.	ref.	-	-	-	-	-	-
	Averse	0.1790	0.0833	-	-	-	-	-	-
CONNECTION TIME	0 mins	-	-	-0.3316	0.0209	-	-	-	-
	5 mins	-	-	-0.4334	0.0024	-	-	-	-
	30 mins	-	-	-0.1374	0.3243	-	-	-	-
	60 mins	-	-	ref.	ref.	-	-	-	-
FREQUENCY OF USE	Never	ref.	ref.	-	-	ref.	ref.	ref.	ref.
	Yearly	0.3759	0.0016	-	-	-0.3465	0.0611	1.0662	0.0027
	Quarterly	0.3378	0.0151	-	-	0.3980	0.0537	0.4343	0.2077
	Monthly	0.3070	0.0405	-	-	-0.1860	0.3552	1.5052	<.0001
	Weekly	0.5253	0.0332	-	-	-0.2462	0.3963	0.5756	0.2007
	Daily	0.5335	0.0192	-	-	-0.4221	0.0671	0.6066	0.1166
CONNECTION MISSED	Yes	0.3800	0.0004	-	-	-	-	-	-
	No	ref.	ref.	-	-	-	-	-	-
APPOINTMENT MISSED	Yes	0.2430	0.0232	-	-	-	-	-	-
	No	ref.	ref.	-	-	-	-	-	-
NUMBER OF OBSERVATIONS		2680		1804		1200		604	
LIKELIHOOD RATIO TEST (p-value)		189.63 (< .0001)		55.39 (< .0001)		41.71 (< .0001)		91.48 (< .0001)	
PROPORTION PREDICTED WITH SUCCESS		65.7%		59.5%		58.6%		74.4%	

Table 7: Choice of a punctuality indicator — Sensitivity analysis

Explanatory Variables		NONE	PERCENT	MEAN	MAX	DaR
AGE	18-24 years	+6.97%	+33.37%	+13.78%	-3.28%	-46.20%
	25-34 years	**0.287**	**0.163**	**0.230**	**0.098**	**0.222**
	35-64 years	+15.33%	+18.45%	+15.09%	-53.02%	-25.60%
	65 years and +	+25.78%	+15.17%	+10.99%	-76.69%	-22.06%
FAMILY SITUATION	Single	**0.328**	**0.199**	**0.240**	**0.088**	**0.145**
	Couple	-2.44%	-5.14%	+9.83%	-48.44%	+25.59%
	Others	+14.94%	-12.08%	+6.52%	-59.28%	+7.83%
INCOME	€0-1000	**0.346**	**0.260**	**0.180**	**0.092**	**0.122**
	€1001-1500	+1.45%	-50.72%	+68.54%	-29.39%	+24.98%
	€1501-2500	+1.73%	-28.70%	+34.91%	-59.30%	+49.29%
	€2501-3500	-7.80%	-31.76%	+57.51%	-69.07%	+56.90%
	€3501-5000	-8.67%	-25.78%	+48.09%	-30.06%	+31.18%
	€5001-6500	-12.72%	-29.04%	+50.64%	-22.93%	+40.40%
	Above €6501	-22.25%	-20.15%	+55.47%	-50.39%	+62.06%
MODE	Motorized vehicles	+6.23%	+6.15%	+2.71%	-10.38%	-16.38%
	Collective transport	-11.53%	+17.39%	-0.30%	+43.64%	-8.43%
	Soft mode	**0.321**	**0.176**	**0.253**	**0.054**	**0.196**
NBM	NBM1 = 1	**0.371**	**0.176**	**0.248**	**0.039**	**0.166**
	NBM2 = 2 OR 3	-32.88%	+22.91%	+9.63%	+136.24%	+2.68%
	NBM3 = 4 TO 6	-60.11%	+35.98%	+26.83%	-4.83%	+57.29%
TT	TT1 $\in$ [1;15]	+101.01%	-24.91%	-15.83%	-54.07%	-27.09%
	TT2 $\in$ [16;30]	+62.81%	-19.38%	-9.26%	-30.27%	-14.56%
	TT3 $\in$ [31;45]	+77.39%	-22.10%	-13.75%	-19.08%	-23.71%
	TT4 $\in$ [46 and +]	**0.199**	**0.231**	**0.287**	**0.079**	**0.204**
COST	COST1 $\in$ [0;1.25]	+60.89%	-20.13%	-10.47%	-43.04%	-26.80%
	COST2 $\in$]1.25; 3]	+36.29%	-20.19%	-7.92%	-19.59%	-4.95%
	COST3 $\in$]3; 5]	+25.00%	-26.12%	-18.53%	+2.06%	+27.51%
	COST4 $\in$]5 and + [	**0.248**	**0.227**	**0.282**	**0.066**	**0.176**
SAFETY MARGIN	SM1 $\in$ [-30; 0[	**0.352**	**0.169**	**0.261**	**0.053**	**0.164**
	SM2 $\in$ [0]	-11.65%	+15.00%	+2.82%	-1.29%	+5.45%
	SM3 $\in$ [1; 10[	+1.14%	+11.55%	-5.04%	-21.09%	+0.54%
	SM4 $\in$ [10; 15]	-13.92%	+11.65%	+0.65%	+19.58%	+10.44%
	SM5 $\in$ [16 and +[	+0.00%	+12.29%	-6.72%	+5.75%	-3.84%
RELIABILITY ATTITUDE	Prone	-17.90%	+7.14%	+11.61%	+45.62%	+7.43%
	Neutral	**0.391**	**0.178**	**0.233**	**0.041**	**0.158**
	Averse	-30.18%	+12.85%	+19.99%	+55.48%	+16.52%
CONNECTION TIME	0 mins	+0.00%	+12.38%	+11.46%	-16.64%	-19.76%
	5 mins	+0.00%	+14.83%	+15.30%	-19.78%	-25.61%
	30 mins	+ 0.00%	+4.95%	+5.38%	-6.42%	-8.94%
	60 mins	**0.327**	**0.176**	**0.239**	**0.062**	**0.197**
FREQUENCY OF USE	Never	**0.462**	**0.148**	**0.214**	**0.052**	**0.125**
	Yearly	-27.06%	+43.01%	+12.96%	-27.40%	+38.28%
	Quarterly	-33.12%	-5.34%	+48.58%	+21.93%	+36.65%
	Monthly	-38.74%	+38.69%	+24.15%	-12.40%	+61.29%
	Weekly	-56.49%	+68.38%	+31.40%	+70.88%	+45.04%
	Daily	-59.74%	+77.01%	+20.56%	+78.42%	+62.27%
CONNECTION MISSED	Yes	-40.94%	+25.90%	+19.87%	+31.44%	+30.74%
	No	**0.381**	**0.174**	**0.241**	**0.050**	**0.154**
APPOINTMENT MISSED	Yes	-36.92%	+24.63%	+17.21%	+ 60.95%	+21.93%
	No	**0.390**	**0.171**	**0.240**	**0.043**	**0.156**

Notes: In bold is the probability of choosing a given indicator; the others figures describe the evolution of the probabilities for any other categories than the reference category, all things being equal.

Considering *Hypothesis 1*, the out option "No indicator is useful to me for planning long-distance train journeys (NONE)" often has the highest frequency, as compared to the other indicators. The proportion of respondents who see no point in having a punctuality indicator is particularly high for respondents who have never used the train transportation (0.462), are risk-neutral on travel time (0.391), have never had bad experiences (missed appointments [0.390] or connections [0.381]), and use only one transportation mode for their regular trips (0.371). These negative results should not obscure the fact that most respondents opt for a punctuality indicator in the choice situations. Nevertheless, MAX seems to be clearly outpaced by the three other indicators PERCENT, MEAN, and DaR. Moreover, among the indicators chosen, PERCENT and MEAN are already proposed on the information websites of the French railway operator and the AQST. DaR, is a new punctuality indicator and, thus, has not yet been suggested by others. Despite its more complex formulation, it appears to be in line with user expectations (probability of choice from 0.122 to 0.222). Finally, the elements identified in the overall analysis (4.1) as influencing the probability of choosing an indicator are confirmed by the econometric analysis. TT and COST have a monotonic effect on the probability: the more time-consuming and costly the trip, the more likely respondents will use a punctuality indicator. Regarding the income variable, the relationship is not strictly monotonous. However, overall, it allows for stating that respondents with higher income are more inclined to use a punctuality indicator. The results support ***Hypothesis 1. Users find retrospective punctuality indicators to be useful for planning their long-distance railway trips***.

The number of modes used for regular trips has a significant impact on the probability that respondents will find punctuality indicators useful for planning long-distance rail travel. Respondents using only one mode of transport for their regular trips prefer the MEAN (0.248), PERCENT (0.176,) and DaR (0.166) indicators for planning long-distance train trips. The probabilities of choosing the three previous indicators are considerably affected by the number of modes used for regular trips. Thus, for respondents using four or more modes, the associated probabilities are respectively 0.315 for MEAN (+26.83%), 0.239 for PERCENT (+35.98%) and 0.261 for DaR (+57.29%). Users' perception of the usefulness of extreme risk indicators is contrasted as follows: while the MAX indicator seems once again be unpopular in planning long-distance railway trips, the usefulness of DaR indicator seems to be confirmed. These results, combined with those of section 4.1, make it possible to validate ***Hypothesis 2. Users using several modes in their regular mobility are more likely to appreciate a retrospective punctuality indicator to plan their long-distance railway trips***.

Reliability-averse respondents (*i.e.*, risk-lovers) find punctuality indicators most useful to plan long-distance trips. However, reliability-prone respondents (*i.e.*, risk-averse) are also more interested in punctuality indicators than risk-neutral respondents. These two categories of people abandon the MAX indicator in favour of MEAN, PERCENT, or DaR. The safety margin, which can be considered as a proxy variable for prudenc,e has an impact on the probability of choosing a punctuality indicator that is not strictly monotonous, depending on the indicators considered. However, the relationship between the safety margin and the perceived usefulness of punctuality indicators by individuals needs to be documented. On the one hand, the more prudent an individual is, the greater his safety margin will be, and, the more he will appreciate having a punctuality indicator, which would only indicate a growing relationship. On the other hand, there is a threshold effect; for small safety margins, the probability of choosing an indicator would increase sharply before decreasing for safety margins to such high levels that the situation would not require the help of a punctuality indicator. This second type of effect is observed for the PERCENT and DaR indicators, whereas nothing compelling appears for the MEAN and MAX indicators (they display alternation in signs that are challenging to interpret). These elements, combined with those in section 4.1, are consistent with the rejection of ***Hypothesis 3. Risk-averse and prudent individuals should be more inclined to use rail punctuality indicators to plan their long-distance railway trips.***

Given the questions in the survey that capture the railway transportation habits of the respondents, we can study their impact on the probability of choosing a punctuality indicator. Perhaps the most striking and obvious fact is that the probability of not finding a punctuality indicator useful to plan a long-distance train trip decreases drastically with the frequency of using a train. This probability varies from 0.462 for a respondent who never uses the train to 0.283 for a monthly train user (-38.74%) and 0.186 for a daily user (-59.74%). Even if the MEAN indicator is at the top of the selected punctuality indicators, the extreme risk DaR indicator is adequately received by daily train users (+62.27%). The MAX and PERCENT indicators also show such evolutions for frequent train users (at least on a weekly or daily basis) but sometimes with alternating signs for other periodicities (which is challenging to interpret) that neither the MEAN nor the DaR show. The results of the econometric model confirm those obtained using the descriptive statistics (4.1) and validate ***Hypothesis 4. Train frequent users are more likely to appreciate having a rail punctuality indicator than occasional users.***

Among variables reflecting railway transportation habits, two items were used to identify possible bad experiences of train users (connections and appointments missed). Thus, missing an appointment or a connection increases the probability of choosing a punctuality indicator as compared to respondents who have never experienced such negative events. These results can be viewed as a sign of a learning effect linked to traveling by train. Once again, the MAX indicator (0.043–0.050) seems to be overshadowed by the MEAN (0.240–0.241), PERCENT (0.171–0.174,) and DaR (0.122–0.156) indicators. Nevertheless, it should be noted that, although the DaR increases its probability very significantly for individuals with bad experiences, it is not sufficient to rank it ahead of the MEAN and PERCENT indicators. Thus, the results partially validate ***Hypothesis 5. Missing an appointment or a connection should encourage users to choose a rail punctuality indicator and, more specifically extreme risk indicators (MAX and DaR).***

4.2.4 Insights, recommendations, and policy implications

The main lesson of our study is undoubtedly that individuals find retrospective rail punctuality indicators useful. Indeed, faced with hypothetical choice situations of a retrospective rail punctuality indicator within the railway trip planning context (with or without connection), in more than 68% of the situations, individuals opted for one of the four indicators proposed.[34]

Beyond the usefulness of retrospective rail punctuality indicators, this then raises the issue of the "good" information to be provided to individuals. The study also provides

[34] This means that the out option was activated in 32% of the situations. A more in-depth study of individuals exercising this out option reveals their specific profile (*e.g.*, risk-neutral, never experienced bad events).

some useful insights for policy makers. While transportation quality agencies or railway operators are accustomed to use indicators such as the percentage of trains on time (PERCENT), the average delay (MEAN), or the maximum delay (MAX), our study shows that the DaR, a totally new punctuality indicator used in the transportation context, can be of interest for users. The DaR, rather based on extreme risks, was particularly well received by frequent train users and/or those who have already experienced negative events (missed connections and/or missed appointments).

The MAX indicator is clearly surpassed by the three other indicators in the minds of respondents as MAX appears to be rather unhelpful for planning long-distance train trips. Of the other three, DaR remains in the third position behind the MEAN and PERCENT indicators. Nevertheless, this is a remarkable performance, since this indicator, derived from the financial literature, is not yet offered on the various information sites, unlike the other two; it is therefore unknown to users. Its formulation is certainly more complex but provides more information to users. It could therefore be useful to propose it as a retrospective indicator of punctuality on traveler information sites, especially if one is interested in behavioral changes of users from the mode choice perspective. For instance, the retrospective punctuality indicators could encourage a modal shift from air toward rail modes.

This example highlights a possible extrapolation of this research to other transportation modes such as air, interurban buses, and urban buses. First, punctuality indicators, and particularly DaR, are easily transposable to any mode of public transport. Stéphan (2015) determined the DaR for air journeys between Paris and Montpellier. Thus, it will be possible to determine whether retrospective punctuality indicators make sense for all public transportation modes, and which punctuality indicators are most appropriate for each transportation mode. It is not clear whether individuals need the same information when considering another transportation mode than rail. Indeed, the degree of punctuality requirements will differ by mode or by trip purposes. Finally, a logical extension of this study would be to study the impact of punctuality indicators in the choice of travel mode. For instance, in a long-distance journey context, we could study the transportation mode choice between car, rail, air, and interurban bus by integrating the more useful retrospective punctuality indicator for each transportation mode like a parameter of choice.

5 Conclusion

This study sheds light on the issue of the relevance of punctuality information to be provided to individuals to assist them in planning long-distance rail trips. The information on punctuality is one element among many in the quality of transport services. In addition to the price variable, the quality dimension should not be neglected if we wish to encourage a modal shift towards modes that generate less negative externalities.

Thus, we surveyed of 670 individuals and subjected them to four choice situations. Each of these choice situations corresponded to planning a long-distance rail trip with or without a connection. In each choice situation, four punctuality indicators were proposed to the individuals, as well as an out option. The respondents were asked to choose one indicator that they felt was relevant to planning the trip described in a specific choice situation. Alternatively, they could select the out option. The punctuality indicators are deliberately based on static or retrospective information and not dynamic or real-time information. Moreover, we collect information on socio-demographic characteristics of respondents, their regular trip behaviors with a special emphasis on risk and prudence attitudes, and railway transportation habits.

Of the four punctuality indicators proposed, three are already offered to users on travel information sites in France (MEAN, PERCENT, MAX), while the fourth is a new and relevant indicator (DaR) derived from the financial literature. The results obtained (descriptive statistics and econometric modeling) show that individuals see some usefulness in these punctuality indicators for planning their long-distance rail trips (67% of choice situations). The MEAN indicator tops the list of user preferences (26% of choice situations), and the PERCENT and DaR indicators follow at 19% and 17% of the choice situations, respectively. The MAX indicator is clearly outpaced (5.5% of choice situations). The usefulness of these punctuality indicators is reinforced by the fact that individuals use several modes for their regular trips, are frequent train users, have already experienced negative events (missed appointments or connections). Risk attitude and prudence of individuals play an important role but not totally in the expected direction.

Finally, this study shows that the DaR, although unknown and more complex in its formulation, exert some characteristics that are appreciated by users. Even so, we believe there is still room for retrospective punctuality indicators alongside dynamic and realtime information to induce user' behavior changes. Future research should attempt to quantify the effect of providing such retrospective punctuality indicators on users' transportation mode choice.

Glossary

AQST	Authority for quality of service in transport
ARAF	Autorité de régulation des activités ferroviaires
ARAFER	Autorité de régulation des activités ferroviaires et routières
ART	Autorité de régulation des transports
CICERO	Center for International Climate Research
DaR	Delay-at-Risk
GCP	Global Carbon Project
HSR	High Speed Rail
IoT	Internet of Things
MAX	Maximum delay
MEAN	Average delay
PERCENT	Percentage of trains on time
SM	Safety margin
SNCF	French national railway operator
VaR	Value-at-Risk

Acknowledgments

The authors thank the three anonymous referees for useful comments that allow to improve the paper. We would like to thank the participants of the 1^{st} Rencontres Francophones Transports Mobilité (Lyon, France, June 2018), as well as the participants of the LAET research team seminar (Lyon, April 2019) for their valuable comments. We are also grateful to the participants of the *Travel Behavior and Public Transport* session of the 15^{th} World Conference on Transport Research (Mumbai, India, May 2019) for the passionate discussions and their constructive remarks. We would like to thank Daniel Serra, Patrice Bougette, and Jimmy Martinez for their careful proofreading. The usual disclaimer applies.

Funding

This research project benefited from the financial support of the French National Research Agency through the program Investments for the Future, under reference number ANR-10-LabX-11-01 (LabEx Entrepreneurship).

References

[1] ARAFER (2018), Le suivi de la qualité de service et des droits des passagers dans le transport ferroviaire de voyageurs en Europe et en France. *L'observatoire des transports et de la mobilité*, Nov. 2018.

[2] Beaud M., Blayac T., Stéphan M. (2016), The impact of travel time variability and travelers' risk attitudes on the values of time and reliability. *Transportation Research Part B*, vol. 93, p. 207–224.

[3] Ben Mansour S., Jouini E., Napp C. (2006), Is There a "Pessimistic" Bias in Individual Beliefs? Evidence from a Simple Survey. *Theory and Decision*, vol. 61, p. 345–362.

[4] Blanton H., Axsom D., McClive K., Price S. (2001), Pessimistic Bias in Comparative Evaluations: A Case of Perceived Vulnerability to the Effects of Negative Life Events. *Personality and Social Psychology Bulletin*, vol. 27, p. 1627–1636.

[5] Börjesson M., Eliasson J. (2011), On the use of average delay as a measure of train reliability. *Transportation Research Part A*, vol. 45, p. 171–184.

[6] Börjesson M., Eliasson J., Franklin J. (2012), Valuations of travel time variability in scheduling versus mean-variance models. *Transportation Research Part B*, vol. 46, p. 855–873.

[7] Brakewood C., Watkins K. (2019), A literature review of the passenger benefits of real-time transit information. *Transport Reviews*, vol. 39, p. 327–356.

[8] Brakewood C., Barbeau S., Watkins K. (2014), An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida. *Transportation Research Part A*, vol. 69, p. 409-422.

[9] BVA (2015), *Observatoire de la vie quotidienne des Français, les Français et les Transports*, avril 2015.

[10] Chang E.C., Asakawa K. (2003), Cultural variations on optimistic and pessimistic bias for self versus a sibling: is there evidence for self-enhancement in the west and for self-criticism in the east when the referent group is specified?. *Journal of Personality and Social Psychology*, vol. 84, p. 569–581.

[11] European Parliament (2007), Regulation (EC) n°1371/2007 of the European Parliament and of the Council of 23 Octobre 2007 on rail passenger's rights and obligations. *Official Journal 315 of 3.12.2007*

[12] Figueres C., Le Quéré C., Mahindra A., Bäte O., Whiteman G., Peters G., Daho G. (2018), Emissions are still rising: ramp up the cuts. *Nature*, vol. 564, p. 27–30.

[13] Fox J. (2016), *Applied Regression Analysis and Generalized Linear Models*, 3^{rd} Edition, Sage Publications.

[14] Friendly M., Meyer D. (2015), *Discrete Data Analysis with R – Visualization and Modeling Techniques for Categorical and Count Data*, 1^{st} Edition, CRC Press, Routledge.

[15] Guirao B., Garcia–Pastor A., Lopez–Lambas M.E. (2016), The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. *Transport Policy*, vol. 49, p. 68–77.

[16] Grechi D., Maggi E. (2018), The importance of punctuality in rail transport investigation on the delay determinants. *European Transport/Trasporti Europei*, Issue 70, Paper n°2, ISSN 1825–3997.

[17] Grotenhuis J.–W., Wiegmans B.W., Rietveld P. (2007), The desired quality of integrated multimodal travel information in public transport: Customer needs for time and effort savings. *Transport Policy*, vol. 14, p. 27–38.

[18] Hull J. (2015), *Options, Futures and Others Derivatives*, 10^{th} Edition.

[19] ITF (2017), Quantifying the Socio-economic Benefits of Transport. *ITF Roundtable Reports*, OECD, Paris.

[20] IRG-Rail (Independant Regulators' Group-Rail) (2018), Sixth Annual Market Monitoring Report, March 2018.

[21] Jackson R.B., Le Quéré C., Andrew R.M., Canadell J.G., Korsbakken J.I., Liu Z., Peters G.P., Zheng B. (2018), Global energy growth is outpacing decarbonization. *Environmental Research Letters*, vol. 13, number 12.

[22] Kahneman D., Tversky A. (1979), Prospect Theory: An analysis of decision under risk. *Econometrica*, vol. 47, p. 263–291.

[23] Kimball M. (1990), Precautionary saving in the small and in the large. *Econometrica*, vol. 58, p. 53–73.

[24] Knight T.E. (1974), An approach to the evaluation of changes in travel unreliability: a safety margin hypothesis. *Transportation*, vol. 3, p. 393–408.

[25] Le Quéré C., Andrew R.M., Friedlingstein P., Sitch S., Hauck J., Pongratz J., Pickers P.A., Korsbakken J.I., Peters G.P., Canadell J.G., Arneth A., Arora V.K., Barbero L., Bastos A., Bopp L., Chevallier F., Chini L.P., Ciais P., Doney S.C., Gkritzalis T., Goll D.S., Harris I., Haverd V., Hoffman F.M., Hoppema M., Houghton R.A., Hurtt G., Ilyina T., Jain A.K., Johannessen T., Jones C.D., Kato E., Keeling R.F., Goldewijk K.K., Landschützer P., Lefèvre N., Lienert S., Liu Z., Lombardozzi D., Metzi N., Munro D.R., Nabel J.E.M.S., Nakaoka S.-I., Neill C., Olsen A., Ono T., Patra P., Peregon A., Peters W., Peylin P., Pfeil B., Pierrot D., Pourlter B., Rehder G., Resplandy L., Robertson E., Rocher M., Rödenbeck C., Schuster U., Scwinger J., Séférian R., Skjelvan I., Steinhoff T., Sutton A., Tans P.P., Tian H., Tilbrook B., Tubiello F.N., van der Laan-Luijkx I.T., van der Werf G., Viovy N., Walker A.P., Wiltshire A.J., Wright R., Zaehle S., Zheng B., (2018), Global Carbon Budget 2018. *Earth System Science Data*, vol. 10, p. 2141–2194.

[26] Lomax T., Schrank D., Turner S. (2003), Selecting travel reliability measures. *Texas Transportation Institute and Cambridge Systematics Report*.

[27] Mbairadjim Moussa A., Stéphan M. (2014), Extreme risk measure for train delay time. *Working paper*.

[28] Mulley C., Clifton G., Balbontin C., Ma I. (2017), Information for travelling: Awareness and usage of the various sources of information available to public transport users in NSW. *Transportation Research Part A*, vol. 101, p. 111–132.

[29] Muntean N., Guizzardi D., Schaaf E., Crippa M., Solazzo E., Olivier J., Vignati E. (2018), CO_2 emissions of all world countries - 2018 Report. *Publications Office of the European Union*, EUR 29433 EN.

[30] Pratt J. (1964), Risk Aversion in the Small and in the Large. *Econometrica*, vol. 32, p. 122–136.

[31] Rietveld P., Bruinsma F.R., Van Vuuren D.J. (2001), Coping with unreliability in public chains: A case study for Netherlands. *Transportation Research Part A*, vol. 35, p. 539–559.

[32] Small K.A., Hsiao C. (1985), Multinomial Logit Specification Tests, *International Economic Review*, Vol. 26, p. 619–627.

[33] SNCF (2012), Perception client des retards et analyse des réseaux transiliens. Annexe 11 du Rapport et documents Valeur du temps, *Evaluation socio-économique du Commissariat Général à la stratégie et à la prospective*, Tome 2, p. 83–86.

[34] Stéphan M. (2015), Fiabilité du temps de transport : Mesures, valorisation monétaire et intégration dans le calcul économique public, Chapitre 1, PhD thesis.

[35] Tam M.L., Lam H.K., Lo H.P. (2008), Modeling air passenger travel behavior on airport ground access mode choices. *Transportmetrica*, vol. 4, p. 135–153.

[36] Tversky A., Kahneman D. (1992), Advances in Prospect Theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, vol. 5, p. 297–323.

[37] Van Lint J., Van Zuylen H., Tu H. (2008), Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A*, vol. 42, p. 258–277.

[38] Van Loon R., Rietveld P., Brons M. (2011), Travel time reliability impacts on railways passenger demand: a revealed preference analysis. *Journal of Transport Geography*, vol. 19, p. 917–925.

[39] Von Neumann J., Morgenstern O. (1944), *Theory of Games and Economic Behavior*, Princeton University Press.

6 Appendix

6.1 Delay-at-Risk (DaR): Definition

We propose a new indicator for rail transportation: Delay-at-Risk noted by DaR. This indicator is derived from the financial extreme risk measures literature, and especially, it is an adaptation of the Value-at-Risk (VaR) in the transportation context.
Suppose that at the time t, we are interested in the risk of a delay for a next arrival $(t+h)$. Let $\mathcal{X}$ be the random variable modeling the delay referred to the risk factor and defined as the difference between the effective and the announced arrival time. We suppose that the train is never early, thus $\mathcal{X}$ is assumed to be a positive continuous random variables with F, a cumulative distribution function (CDF), and Q, a quantile function.

Definition

For a risk level $\alpha \in [0, 1]$, the Delay-at-Risk at level α, for time horizon h, noted $DaR_\alpha(h)$, is delay time such that:

$$\mathbb{P}\left[\mathcal{X} \leq DaR_\alpha(h)\right] = \alpha, \tag{3}$$

or alternatively

$$DaR_\alpha(h) = F^{-1}\left(1 - \alpha\right), \tag{4}$$

where $F^{-1}\left(\cdot\right)$ is the pseudo-inverse of the CDF.

From this definition, the probability that the delay time at ahead horizon h would be greater than or equal to $DaR_\alpha(h)$ over the next time horizon is α. An alternative interpretation of this risk measure is that, with probability $1 - p$, the delay over the time horizon t is less than DaR_α. This definition also shows that the DaR is concerned with the upper tail behavior of the delay CDF at time horizon h. When the $CDF(h)$ is known and $\mathcal{X}_{t+h}$ is a continuous random variable, Equation (4) expresses $DaR_\alpha(h)$ as the $(1-\alpha)$ quantile of the $CDF(h)$. For instance, a $DaR_{0.05} = 30$ minutes implies that the maximum delay time is 30 minutes with a probability 5%, or equivalently, 95% of journey will have a maximum delay time of 30 minutes.

6.2 Survey Questionnaire

The survey has been proposed in French. We propose a translation of original survey in English.

We are a group of researchers at Rennes 1 and Montpellier Universities. Our research works are organizing on the theme of travel time reliability.

The survey deals with information type that you will wish have in order to organize your trip. We are asking you to answer all questions, even if some questions seem take you away from initial topic.

There are no right or wrong answers. The most important is that your answers match your feelings, your preferences or your wishes.

To answer our survey, you need around 10 minutes. Thank you for your participation and the help given in realization of our research works.

1 - **Part 1: Socioeconomics and demographics characteristics**

1. You are:
   - ☐ A woman
   - ☐ A man
2. What is your age bracket?
   - ☐ 15-18 years
   - ☐ 19-25 years
   - ☐ 26-35 years
   - ☐ 36-50 years
   - ☐ 51-65 years
   - ☐ 66-75 years
   - ☐ More 75 years
3. Now, you are:
   - ☐ Single
   - ☐ In couple
   - ☐ Widowed, Divorced.
4. Do you have children?
   - ☐ Yes
   - ☐ No
5. What is your household composition?
   - Number of individual that have 14 years old and more, you included:
   - Number of individual that have less 14 years old:
6. What is your household monthly net income?
   - ☐ 0-500 €
   - ☐ 501-1 000 €
   - ☐ 1 001-1 500 €
   - ☐ 1 501-2 500 €
   - ☐ 2 501-3 500 €
   - ☐ 3 501-4 500 €
   - ☐ 4 501-6 000 €
   - ☐ 6 001-7 500 €
   - ☐ 7 501 € and more

2 - **Part 2: Your regular trip behavior**

7. In the context of your regular trips (work, study, leisure, etc.), what is the transportation mode you use frequently or on the longer part of your trip? (Only one possible response)

   ☐ Car
   ☐ Train
   ☐ Bus
   ☐ Tramway
   ☐ Carpool
   ☐ Biking
   ☐ Motorbike
   ☐ Walking
   ☐ Others - Specify:

8. In the same previous context, do you use a combination of several transportation modes (e.g. Train + Biking)?

   ☐ No
   ☐ Yes – Which?

9. For your regular trips,

   - What time do you leave your home (or other localisation)?
   - What time do you start your activity at destination?
   - What is your average travel time of your trip (in minutes) ($\bar{X}$)?

10. In your opinion, what is the monetary cost (in euro) of your previous trip?

11. During your regular trips, do you have delay problems?

    ☐ No, never
    ☐ Yes, but rarely
    ☐ Yes, sometimes
    ☐ Yes, frequently
    ☐ Yes, very frequently

12. Now, we are presenting you 2 situations related to your regular trips, choose the one you prefer:

Situation 1: You have a 50% chance that your travel time will be $\bar{X} - \frac{1}{2}\bar{X}$ minutes, and a 50% chance that your travel time will be $\bar{X} + \frac{1}{2}\bar{X}$ minutes.

Situation 2: Your travel time is $\bar{X}$.

☐ I prefer situation 1.
☐ I prefer situation 2.

☐ I am indifferent between both situations.

3 - **Part 3: Reliability Information**

13. You must book train ticket for an unconnected train. The railway operator announces a travel time of 3h20 mins. The booking website gives the following indicators:

    ☐ Indicator 1: The rate of trains that are on time is 74%.

    ☐ Indicator 2: The average delay is 5 mins for all trains, and is 27 mins for all late trains.

    ☐ Indicator 3: In 5% of cases, the train will have a delay higher than 23 mins.

    ☐ Indicator 4: The maximum delay is 192 mins.

    ☐ Indicator 5: No information is useful.

    Please, choose the best indicator for you.

14. You must book train ticket for a trip with a connection. The first part of trip corresponds with an announced travel time of 3h20 mins by railway operator. The connection time with the used transportation mode for the second part of trip is estimated to 5 mins. The booking website gives the following indicators:

    ☐ Indicator 1: The rate of trains that are on time is 74%.

    ☐ Indicator 2: The average delay is 5 mins for all trains, and is 27 mins for all late trains.

    ☐ Indicator 3: In 5% of cases, the train will have a delay higher than 23 mins.

    ☐ Indicator 4: The maximum delay is 192 mins.

    ☐ Indicator 5: No information is useful.

    Please, choose the best indicator for you.

15. Same question with a connection time estimated to 30 mins.

    ☐ Indicator 1: The rate of trains that are on time is 74%.

    ☐ Indicator 2: The average delay is 5 mins for all trains, and is 27 mins for all late trains.

    ☐ Indicator 3: In 5% of cases, the train will have a delay higher than 23 mins.

    ☐ Indicator 4: The maximum delay is 192 mins.

    ☐ Indicator 5: No information is useful.

    Please, choose the best indicator for you.

16. Same question with a connection time estimated to 60 mins.

    ☐ Indicator 1: The rate of trains that are on time is 74%.

    ☐ Indicator 2: The average delay is 5 mins for all trains, and is 27 mins for all late trains.

- ☐ Indicator 3: In 5% of cases, the train will have a delay higher than 23 mins.
- ☐ Indicator 4: The maximum delay is 192 mins.
- ☐ Indicator 5: No information is useful.

Please, choose the best indicator for you.

17. How often do you use the train?
    - ☐ Several times a week
    - ☐ Once a week
    - ☐ Once a month
    - ☐ Once a quarter
    - ☐ Once a year
    - ☐ Never
18. Have you ever missed a connection because of a train delay?
    - ☐ Yes
    - ☐ No
19. Have you ever been late because of a train delay?
    - ☐ Yes
    - ☐ No
20. Did you have any difficulties to answer to this questionnaire?
    - ☐ Yes
    - ☐ No
    - If Yes, which ones ?