

This is the unformatted version of: Gao, Y. and Schmöcker, J.-D. (2021). Modelling sequential ticket booking choices during Chinese New Year. *Transportation*, 48(4), 1987-2010. <https://doi.org/10.1007/s11116-020-10118-8>.

Modelling sequential ticket booking choices during Chinese New Year

Yuhan Gao and Jan-Dirk Schmöcker
Department of Urban Management, Kyoto University
{gao,schmoecker}@trans.kuciv.kyoto-u.ac.jp

Abstract

Transport systems in China face extreme capacity shortages during the Spring Festival travel season. This study therefore explores traveler's decision making behavior when booking tickets during this season for a specific origin destination relation. Data are obtained from a mixed RP/SP survey that reflects the ticketing policy in China. Loop questions are programmed inside the questionnaire to investigate changes in behavior after people experienced failure to obtain their preferred ticket. Hybrid discrete choice models are built based on the survey data to reflect the first two levels of an individual's choice sequence. The first level reflects respondents' preferred choice. For the second level, the best alternative choice, a generalized choice set is introduced that is able to simplify the complex alternatives. Two simulations are conducted based on the estimation results which are evaluated based on the number of tickets that could be obtained by the generated population. The simulation results suggest that under fixed total railway capacity, the over-expansion of HSR can have an overall negative impact, especially for the low-income group and we show the possibility of creating more capacity for the low-income group by staggered ticket sales. We suggest the methodology introduced in this study can support the discussion regarding socially optimal ticketing policies under severe capacity shortages.

Keywords: High Speed Rail, Ticket shortage, Sequential choice, Chinese New Year

1. Introduction

1.1 Background and motivation

The Spring Festival travel season is a time period of extremely high travel volume around the Chinese New Year. Chinese return home for reunion with families on the Lunar New Year's Eve, following a long-held tradition. The travel season begins 15 days before the Chinese New Year's Day and lasts for about 40 days. The departure peak usually is three to four days before the Lunar New Year's Eve, and the return peak is usually the last few days of the 7-day national holiday following the Lunar New Year's Day. Many people travel from their study or working places to their native place in this period. Most of the trips are long distance, between the coastal urban areas to the inland rural areas. It is reported that passenger trips of every Spring Festival travel season have exceeded the population of mainland China in the past decade. In the 2017 Spring Festival travel season, 2.98 billion trips were generated nationwide, 2.29% more than in 2016 (State Council of the People's Republic of China, 2017). The current transportation systems in China cannot fully meet this extremely large demand, thus a series of traffic as well as social problems are emerging.

The insufficient capacity is particularly reflected in rail ticket shortage. The railway, including high-speed railway (HSR) and the standard service, is the most popular option for long distance travel in the Spring Festival, because of its affordable price, convenience and network coverage. All long-distance train tickets are personalized and non-transferable to avoid black-market prices, and further only a strict contingent of tickets for each origin-destination pair is sold. The fares depend on ticket OD, class of train (HSR or normal railway), and class of seat. (We ignore "luxury choices" such as HSR business class tickets in the following in order to simply the choice set and due to their low market share). The tickets can be bought through internet or telephone 30 days before the travel day. Popular routes are often sold out within seconds of their release since. Therefore people often wait online at exactly the time when tickets for their preferred date become available.

In case one is not successful in buying a direct ticket, often alternative plans are taken up

instead of giving up traveling. Some people aim to, for example, buy tickets to a destination which is further downstream on the train line of their true destination but get off earlier at their true destination. Also to be noted is that if people fail to obtain train tickets initially, they can hope for obtaining a small amount of additional or returned tickets that are sold at a later date.

Besides the high demand for rail travel during Spring Festival, aviation is for many journeys very competitive given the size of China which means that - even with HSR – rail journeys can take a considerable amount of time. The pricing in the domestic aviation market is largely market driven, with some government regulation. Flight ticket prices rise sharply during the Spring Festival season, usually remaining at the maximum limit set by the government, and because of their high price often some tickets remain available even for popular routes until nearly the travel date.

With this background, this study therefore explores people's decision making behavior when buying tickets for long distance travel during Chinese New Year travel season, and how their perceived chance of getting a ticket changes, especially when their preferred traffic mode becomes unavailable. Due to the scale of the travel generated during Chinese New Year, we suggest it is worth studying this problem by itself. Furthermore, the above described features suggest that the Spring Festival rush is a suitable case also for studying in more general decision making in transportation networks with extreme capacity shortage. Our objective is to understand people's preferences and their strategy to cope with failure in obtaining their preferred ticket. Given these strategies we aim to show the effect of alternative ticketing policies.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature. Section 3 describes our survey and the proposed model to reflect the decision making process. Sections 4 and 5 then describe the resulting choice models for preferred and second preferred choice respectively. Section 6 utilizes these results to construct a simulation model for different ticketing policies. Conclusions are drawn in Section 7.

2. Literature review

2.1 Spring Festival rush

There is significant literature (mostly in Chinese) on the social problems resulting from the insufficient capacity during Spring Festival. The price setting and regulation in China railway has evolved since the 1950s. With the development of Chinese society, Scales et al (2011) note that the fixed price system seems to be out of date. They suggest some alternatives that aim at changing the old price regulation gradually, such as piloting market-based pricing for some routes. There is also a long lasting debate whether railway ticket prices during the Spring Festival should increase or not. Zhou and Kang (2006) suggest that in order to protect consumers' benefit from ticket scalper as well as to reduce the waste of social resources, it is necessary to establish a pricing mechanism controlling particularly for the Spring Festival travel season. More specifically, they suggest adopting Dutch-price auctioning for calculating reasonable prices. Zhou (2013) conducted a comparison between air and railway ticketing policies in China from the point of view of law and economics and proposes to adopt floating pricing by different time periods to reduce the social problems related to ticket shortages.

Besides pricing, there is ample sociological research on the Spring Festival rush. Shao (2008) focused on the behavior of migrant workers and suggested "sociological treatment" is required such as educational policies to promote equity instead of merely enhancing transportation capacity. For this it is therefore important to understand which population groups obtain what type of tickets and which groups are the likely victims of ticket shortages which is one objective of our study.

2.2 Ticket booking behavior and risky, sequential decisions

There is some limited research on generally understanding rail choices in China. Chen et al. (2007) established a double-nested logit model to analyze the passenger's choice behavior in the Chinese railway system using stated preference data for the Shijiazhuang-Taiyuan Railway and found that departure time and personal social economic factors, such as monthly income, have effects on travel mode choice. Further, transport ticket purchase behavior has

been explored mainly with a view of market segregation and for appropriate ticket pricing (e.g. Wang and Huang, 2016; Cheng and Huang, 2014). In this research, however, the focus is on the sequence of decisions when deciding on aiming for a specific ticket and possibly agreeing to purchase a ticket that is not one's preferred choice.

There is a substantial body of literature established on such "sequential choice modelling" over the last decades. Whereas in early papers decisions were modeled independently subsequent choice literature demonstrates that correlation in the decisions of the same person for the "upper" and "lower" problem should be considered. Fu and Wilmot (2004) for example develop a sequential logit model with binary choices regarding whether to evacuate for a hurricane and estimate the model by considering all binary choices simultaneously instead of estimating multiple models along time intervals. Furthermore, in a range of problems, the expectation of the outcome of subsequent decisions might influence one's attitude and with it one's decision making in the first decision. Eckstein and Wolpin (1989) and Puterman (1990) summarize models regarding dynamic decision-making. Instead of using only the current alternative-specific utilities for interpreting choice probabilities, value functions are introduced in order to take account of future utilities given current condition.

For literature that captures these effects in other choice contexts we mention Fosgerau et al (2013) who introduce recursive logit in modelling individuals' sequential route choice with not only instantaneous utility under given condition but downstream utility expected being taken into account. Further Liu and Cirillo (2017) develop dynamic discrete choice models to model and forecast car purchase behavior and preferences in a finite time horizon setting. To some degree closer to our problem is Urata and Hato's (2013) work who formulated the trade-off between delay penalties and engaging immediately in evacuation activities after a disaster occurred based on a logit model. Assuming decision makers have insufficient information for future decision-making states, Oyama and Hato (2017) further introduce a discount factor into expected downstream utility in a recursive logit approach for modelling people's myopic behavior in sequential route choices.

In this study we suggest, however, that the problem can be simplified. The “investment cost” in terms of time and money to aim for one’s preferred, risky choice is minor. The cost is mainly to possibly forego purchasing other tickets at exactly the same time. However, in practice and in our survey set up there is one highly preferred option with limited capacity but some other options usually have some spare capacity. Therefore, we in fact observe that people usually aim for the option that is perceived more risky in first place. In summary, we suggest that correlation in behavior in the different choice levels needs to be considered but that travelers do not exhibit “strategic” behavior for their first choice, but initially always aim for their preferred option. If the person fails s/he is then faced with a choice of going for a less preferred alternative or waiting for the next day when different tickets become available. We conclude by noting that for the same reasons of a person aiming at first “anyway” for his preferred choice, we suggest that for our problem also utility maximization appears to be suitable. Regret minimization for decisions under uncertainty as in for example Quiggin (1994) or Chorus et al (2008) would again presume that the risks of failing or to encounter delays are evaluated within one’s first choice.

3. Decision Making Process and Resulting Survey

3.1 Decision making process and survey implementation

Based on the above described ticketing policy for the Spring Festival travel season we depict the choice process of an individual in below flow diagram (Figure 1). The figure describes people’s decision making process within each day and from day to day when aiming to buy tickets. t is 30 days before each travel day when train tickets become available; U_{nkt} is the utility of option k on day t for person n ; W_{nt} is the utility of not buying a ticket on day t but waiting instead for tickets on day $t+1$; superscript (o) denotes the person specific rank of each option k in descending order. While planning travel for the Spring Festival, individuals determine the acceptable departure time period, rank their preference and then attempt to purchase a ticket. If the tickets of the first choice are sold out, individuals may change to an

alternative, lower ranked plan, or decide to stick to their favorite travel mode and wait for another set of tickets becoming available on the next day. Individuals make decisions on each day when tickets of their acceptable travel dates are released.

An online survey is created with a scenario as in Figure 2 which allows the respondents to project themselves into the situation of booking tickets for long-distance travel in the ticketing peak of the Spring Festival. We presume an origin (Guangzhou) and destination (Hefei) where both rail and flight are likely attractive but bus not due to the large distance. To increase understandability of the options and the difficulty with obtaining tickets for the given route, respondents are recruited only from around Guangzhou and Hefei. We pick four kinds of ticket which are the major options in our scenario (as well as many other long distance origin destination pairs inside China). Details are summarized in Table 1.

Table 1: Ticket option and policy in the survey

	NORMAL RAILWAY		HIGH-SPEED RAILWAY	AVIATION
TYPE/CLASS OF TICKET (HEREINAFTER ABBREVIATED AS)	Seat ticket (Seat)	Sleeper ticket (Sleeper)	Regular class (HSR)	Economic class (Flight)
TICKET RELEASE DATE	30 days before departure day			No constraint
PRICE	Fixed			Rises sharply, remaining at the maximum limit. Simplified as fixed

The respondents are asked to answer the questionnaire considering their current living circumstances, including family and financial situation. In addition, loop questions in line with the “next-day choices” are used to investigate how people change their behavior after experiencing failure in ticket booking. Respondents are firstly asked about their perceived risk to get a ticket of their favorite mode. Next, the respondents are told that they have failed to get

a ticket for their chosen favorite mode. A follow-on question asks about what alternative plans they would try, including changing mode or taking transfer routes as well as waiting for tickets released on the next day. The alternatives need to be ranked.

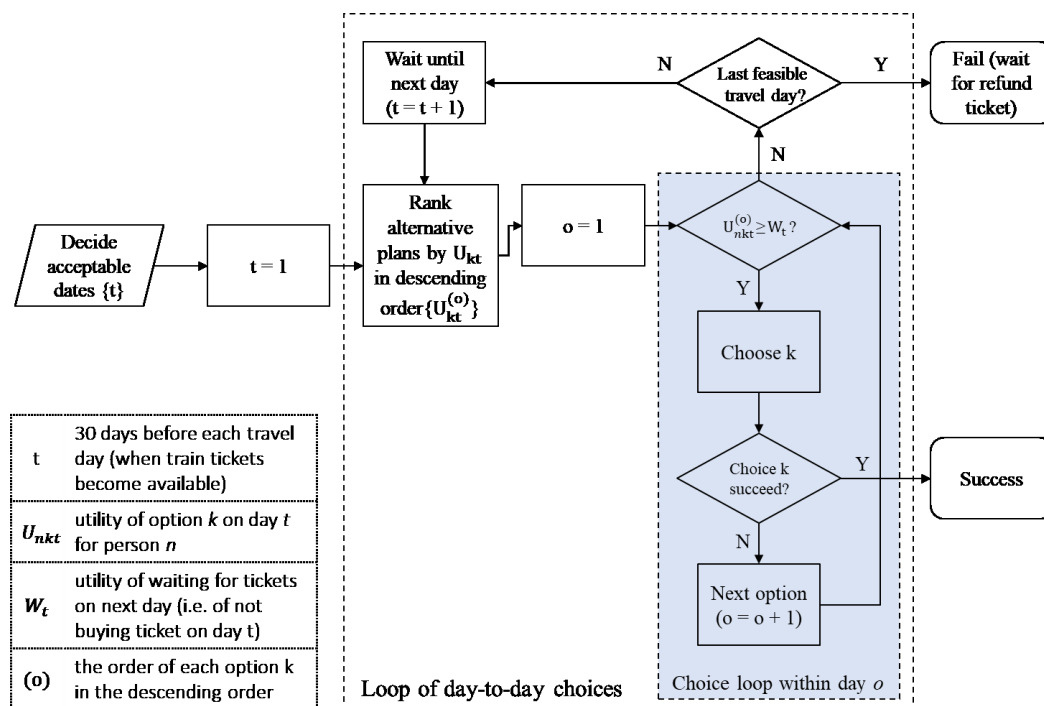


Figure 1: Proposed choice sequence for buying tickets during spring festival travel season



Figure 2: The scenario given to respondents in the survey

The survey is conducted between August and September 2017. The questionnaire is distributed by widely used SNS applications in China which allows us to reach various population groups. Since our target is the choice behavior in the final week before the Lunar

New Year's Eve, the survey distribution in particular tried to avoid SNS networks that would attract too many high school and university students since students start their Spring Festival vacation long before the last week of lunar year.

3.2 Survey statistics

Table 2 shows some descriptive statistics of the 452 completed surveys and Table 3 a comparison of the mean and median of annual household income from the survey sample and the income statistics of the Chinese Household Income Project 2013 (CHIP 2013). Our sample is fairly young which is partly due to the data collection method. We further observe that two out of three respondents has an experience of travelling during the spring festival in the last five years. Regarding income, considering the economic growth in China, the mean and median from CHIP is adjusted by nominal GDP growth rate of China. Both the mean and median of the annual household income from our survey sample are significantly higher than the mean and median from the published data. The bias is expected since insufficient samples of low-income people, such as migrant workers could be obtained. This bias has been considered in the following modeling and simulation. We observe that 7% of the respondents refuse to answer their annual household income. Sanko et al (2014) and Bahamonde-Birke and Hannapi (2016) proposed latent variable treatment for correcting the measurement errors by non-reporters. Unfortunately, we do not have enough information to correct the bias in such a way. Among the information that we have, using bivariate correlation, we find that no individual socio-demographic attribute is highly correlated with refusing to answer the income question. We do find, however, that those choosing not to answer the income question choose HSR, but not Flight, more often than the two standard rail options.

Table 2: Aggregate survey statistics (mean and standard deviation)

	Number of observations	Mean	Standard deviation
Age	452	36.77	11.47
Gender(female)	452	0.58	0.49
Spring Festival rush experience in the last 5 years (binary)	452	0.67	0.47
Income dummy (answered)	452	0.93	0.26

Table 3: Mean and median of annual household income

Annual household income (10,000 CNY)	Sample (2017)	CHIP (2013)	
		2013	Adj.
Mean	14.10	5.53	7.70
Median	8.00	4.22	5.87

Respondents' mode choice is presented in Figure 3. Among all the valid responses, HSR is the most popular travel mode in the presumed scenario. 271 of 452 respondents chose HSR as their first choice travel mode. Further, 132 respondents said they would choose HSR if they cannot obtain their first (preferred) mode tickets. Only considering the first choice, flight is the second-most popular mode among the respondents. However, it is unpopular as a second choice, whereas Seat, Sleeper and HSR are almost evenly matched. We suggest this finding is partly due to above discussed sample bias. With a sample better reflecting the population characteristics, Seat options would be more attractive due to the low income of the under-represented migrant workers. Regardless of this, we suggest that for our choice model the sample is useful as no choice seems to dominate the others too significantly.

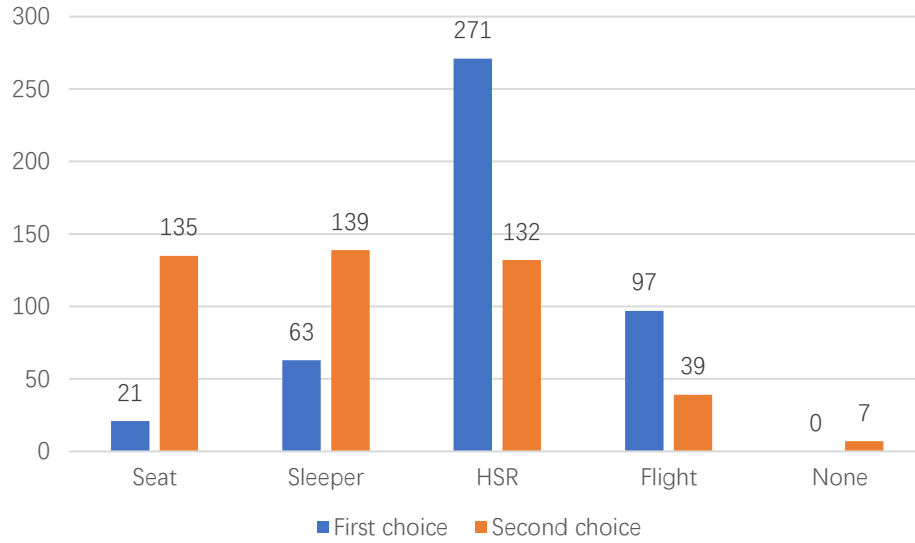


Figure 3: Respondents' mode choice

3.3 Choice sequences

Some of the respondents answered they would try many if not all options before choosing “Wait (for next day)”. Ten respondents would aim for all seven options before choosing to wait for the next day, meaning they would attempt all four ticket types as well as rail options for Seat, Sleeper and HSR that include transfers. The exception are respondents in the Seat group who often chose Wait after trying only few alternatives. Only four out of 21 respondents who answered Seat as their preferred ticket, would try to book an HSR or flight ticket. This observation confirms that the choices of the Seat group may be limited by their income.

We illustrate these findings by the dominant choice sequences in the first three choice levels presented in Figure 4. The alternatives with prime refer to alternatives without changing the mode, for example, in Figure 2, to buy a ticket Guangzhou to Nanjing (and get-off in Hefei) which is obviously more expensive. The diagram confirms that HSR is also a very popular alternative in the second choice level. All the three groups are attracted by HSR, except for the Seat group, which seems to be a separated market. The choice of Wait dominates the third level for all modes. Looking specifically at the Wait-ratio by each level, Figure 5 shows that most respondents prefer to wait for the next day after experiencing two to three booking failures.

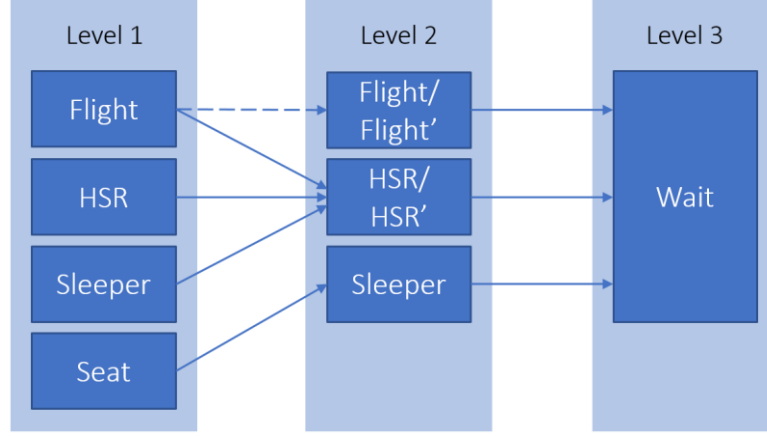


Figure 4: Dominant choice sequences

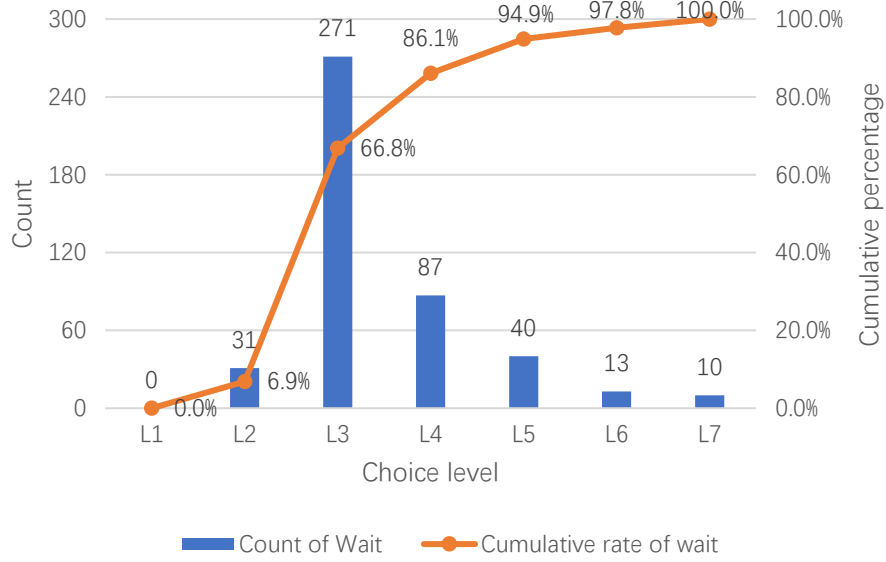


Figure 5: Number of people who choose to wait by decision levels (*L1 indicates preferred choice, L2 choice if L1 is not available, L3 choice if L1 and L2 are not available and so on.*)

4. Multinomial logit model with latent variables for first choice

4.1 Choice set and Model formulation

In the following we focus on the first two levels of people's choice. A separated two-level

sequential model is introduced. Given our observation on the general apparent attractiveness of HSR as well as the different type of travel experience a Sleeper journey offers, the decision making for the first choice is formulated by a multinomial logit model combined with latent variables modelling (Ben-Akiva et al, 2002). For individual n , the choice sets are denoted by C_n^1 and C_n^2 , where the superscript represents the choice level. In this section, the first level of the sequential model is introduced. Initially we presume an identical choice for all respondents, that is:

$$C_n^1 = \{Seat, Sleeper, HSR, Flight\}. \quad (1)$$

We assume individual's preference towards service quality effects choice for the first mode. In the survey respondents are asked to evaluate how important the time cost, comfort, punctuality and flexibility is on a 5-point scale with endpoints “not at all important” and “extremely important”. The resulting subjective data are used as indicators for a latent variable identified as *Service Quality Preference*. We note that defining a latent variable with indicators too close to the attributes of the choices itself creates problems as to how the results can be interpreted as discussed in Chorus and Kroesen (2014). To some degree with face this problem here, but we note that relationship between indicators and choices is not always clear. For example, it is hard to argue whether Flight or HSR is more comfortable or on-time. Further, punctuality and flexibility will be the same for Seat and Sleeper.

We further use income as a direct explanatory variable. In the China Household Finance Survey 2013 which we referred to in Table 3, household income is measured on a 7 point scale, with ranges: 1) 0-8,000 CNY (*extremely low income group*); 2) 8,000-29,999 CNY (*low income group*); 3) 30,000-59,999 CNY (*middle-low income group*); 4) 60,000-99,999 CNY (*middle income group*); 5) 100,000-199,999 CNY (*middle-high income group*); 6) 200,000-299,999 CNY (*high income group*); 7) more than 300,000 CNY (*extremely high income group*). To acknowledge that income effects are nonlinear in the model we also use group dummies, but after testing, merge these into three categories, low income (group 1-2), middle income (group

3-5) and high income (group 6-7). An income dummy is defined to control for the “do not want to answer” option in the annual household income question (equal to 1 if answered and 0 otherwise).

For our latent construct *Service Quality Preference* the structural equation is characterized in linear form. To control for sociodemographic effects, specific dummy variables of *gender (female)*, *student*, and *migrant worker* are introduced into the structural function of the latent variable. Ordered logit is used for modelling the measurement equations of the indicators.

Figure 6 shows the model framework. We note the absence of time and cost as our study focuses on a single origin destination relation. Thus we cannot control for the variation in evaluation of these attributes across the sample as well as with respect to the specific alternatives. This will lead to an upward bias in heteroscedasticity in the sample that we aim to control for with the other observed and latent variables discussed in the following.

Latent variables are represented as oval, observed variables are represented as rectangles, and indicators are represented as rounded rectangles. Solid arrows represent structural relationships and dashed arrows represent measurement relationships. All the models are estimated by maximum likelihood estimation following the procedure in PadasBiogeme (Bierlaire, 2018). The estimation results are presented in Table 4. We utilise normalization as in Daly et al. (2012) where the variance of each measurement equation is fixed to 1 and the first threshold is fixed to 0, while the variance of the structural equation as well as the intercept of each measurement equation are left free. For further details on the model formulation of latent choice models we refer to Ben Akiva (2002) and Bierlaire (2018).

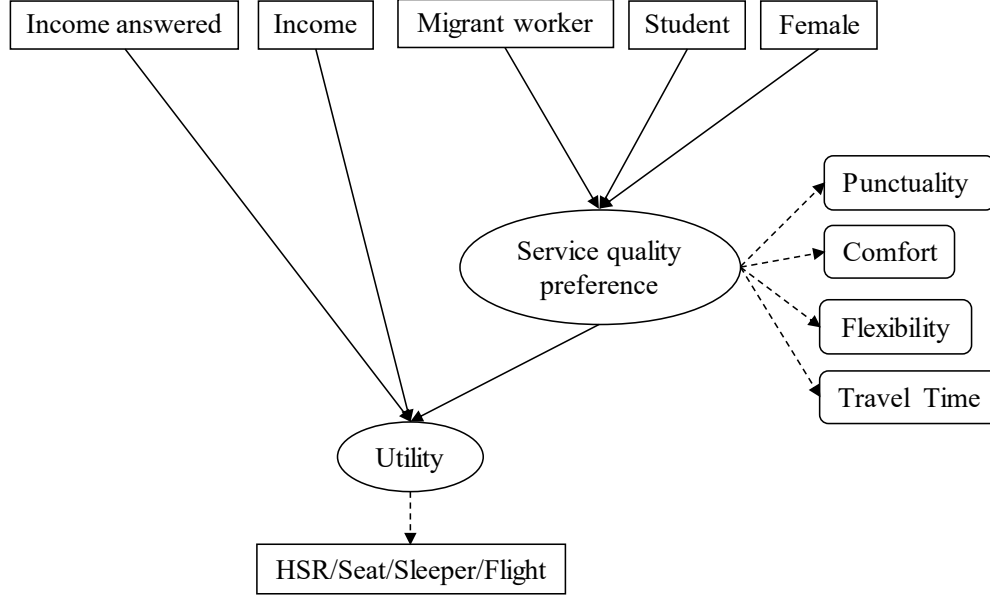


Figure 6: Framework of the Latent MNL model for Preferred Choice (Level 1)

4.2 Result discussion

Our model fit appears to be acceptable as we obtain an Adjusted rho-square ($\bar{\rho}^2$) of 0.477. The alternative specific constants (ASCs) of Seat, Sleeper and Flight are all negative with respect to HSR showing a general preference for this mode. Since, as explained, we can not obtain (mode specific) values of time, cost and comfort from our study, the ASCs must be interpreted as indicating that HSR appears to have the most preferable trade-off between these attributes and that Seat is, all else being equal, most disliked, probably due to its low comfort. (Note that Seat and Sleeper have the same travel time).

Table 4: Level 1 model estimation results

	HSR	SEAT	SLEEPER	FLIGHT
	Estimation t-test	Estimation t-test	Estimation t-test	Estimation t-test
Alternative specific constant	ref. /	-3.18 (-2.92***)	-1.87 (-2.99***)	-1.56 (-3.24***)
Middle income	ref. /	-2 (-3.39***)	0.227 (0.624)	0.412 (1.2)
High income	ref. /	-3.24 (-2.88***)	-1.05 (-1.95**)	0.767 (2.04**)
Income dummy (answered)	ref. /	2.2 (1.95*)	0.746 (1.06)	-0.321 (-0.575)
Latent Variables				
Service Quality Preference	ref. /	-1.06 (-3.33***)	-0.414 (-2.68***)	0.393 (2.75***)
Structural Equations for Latent Variables				
	estimation	t-test		
Female	0.842	3.49***		
Migrant Worker	-1.52	-2.6***		
Student	-1.1	-2.27**		
σ	1.09	5.48***		
Measurement Equations for Latent Variables				
	estimation	t-test		
intercept	5.02	9.84***		
λ	1	/		
τ_1	1	/		
δ_1	2.51	6.06***		
δ_2	1.99	9.28***		
δ_3	2.75	12.9***		
intercept	5.27	10.7***		
λ	0.953	4.02***		
τ_1	1	/		
δ_1	2.18	5.00***		
δ_2	3.17	11.4***		
δ_3	2.65	12.6***		
intercept	4.69	11.4***		
λ	0.836	4.29***		
τ_1	1	/		
δ_1	2.31	6.3***		
δ_2	2.18	10.8***		
δ_3	2.97	13.8***		
intercept	4.81	11.2***		
λ	0.689	4.17***		
τ_1	1	/		
δ_1	1.83	4.94***		
δ_2	1.87	8.8***		
δ_3	2.7	14.2***		
Model Statistics				
Sample Size	452	Likelihood ratio test	4788.933	
Initial log likelihood	-4941.684	Rho-square	0.485	
Final log likelihood	-2547.218	Adjusted Rho-square	0.477	

* Significant at 10%; ** Significant at 5%; *** Significant at 1%

We find *Service Quality Preference* significantly effects people's decision making with the expected signs. We obtain a positive sign for flight but negative signs for Sleeper and Seat. Therefore those who put more emphasis on a high quality service are also more likely to choose the faster, expensive options. Looking at the structural equation of *Service Quality Preference*, estimation results suggest that females care more about service quality than male. It seems that migrant workers and (university) students, who are most likely to have the shortest time with families in a year over other occupation groups, care relatively less about the service quality. Among all the indicators, those who care about service quality are likely to rate the importance of punctuality highest, followed by travel time, flexibility and then comfort.

Even controlling for service quality preference, we observe a significant effect of income with the expected signs: *High income* has a positive effect on preference for flight, a negative effect for Sleeper, and a strong negative effect for Seat. For *Middle income* we cannot find significant effects on Sleeper and Flight, while there is still a negative effect for Seat. We note that those who answered the income question tend to choose the cheaper options suggesting that - despite above discussed insignificant bivariate correlations of the dummy with income categories – controlling for other variables, possibly respondents with higher income tended to not to state their income.

5. Choice model if preferred option is sold out

5.1 Choice set simplification

For the choice set C_n^2 we include the option “Wait”. Furthermore, the choice set is now person specific. For one, the preferred choice is not available anymore and furthermore we include variations of the same mode that was chosen in the first choice. However, to allow robust modelling, we simplify the choice as described in the following and shown in Figure 7.

There are only 4 alternatives in the generalized choice set, which are “Up”, “Down”, “Stay” and “Wait”. A choice of “Up” represents the choice of any other mode that is more expensive than the one chosen in Level 1. Similarly, a choice of “Down” is defined as choosing a mode that costs less (and is less convenient) than one’s preferred choice. The without-changing-mode alternatives are classified as choices of “Stay”, which refers to (less convenient and more expensive) variations of the preferred choice. The definition of choice “Wait” is unchanged.

With this let C^2 denote the generalized choice set for Level 2

$$C^2 = \{Up, Down, Stay, Wait\}. \quad (2)$$

Respondents in the (preferred choice) Seat group cannot go further “Down”, and similarly persons who chose Flight as their preferred option can not go further “Up”. This leads to Equations (3) to (5) where ***Se***, ***Sl***, ***H***, ***F*** denotes the group of persons who choose Seat, Sleeper, HSR or Flight as their preferred Level 1 choice respectively.

$$C_n^2 = \{Up, Stay, Wait\}, \forall n \in \mathbf{Se}. \quad (3)$$

$$C_n^2 = \{Down, Stay, Wait\}, \forall n \in \mathbf{F}. \quad (4)$$

$$C_n^2 = \{Up, Down, Stay, Wait\}, \forall n \in \mathbf{Sl}, \mathbf{H} \quad (5)$$

Figure 7 shows the resulting conversion of the decision tree. Due to above simplification we can utilize an MNL approach, controlling for available options and utilizing partially different explanatory variables. We note that nested and cross-nested logit models were also tested, but we found that the (still) large number of options led to less interpretable results and we prefer the model reported in the following.

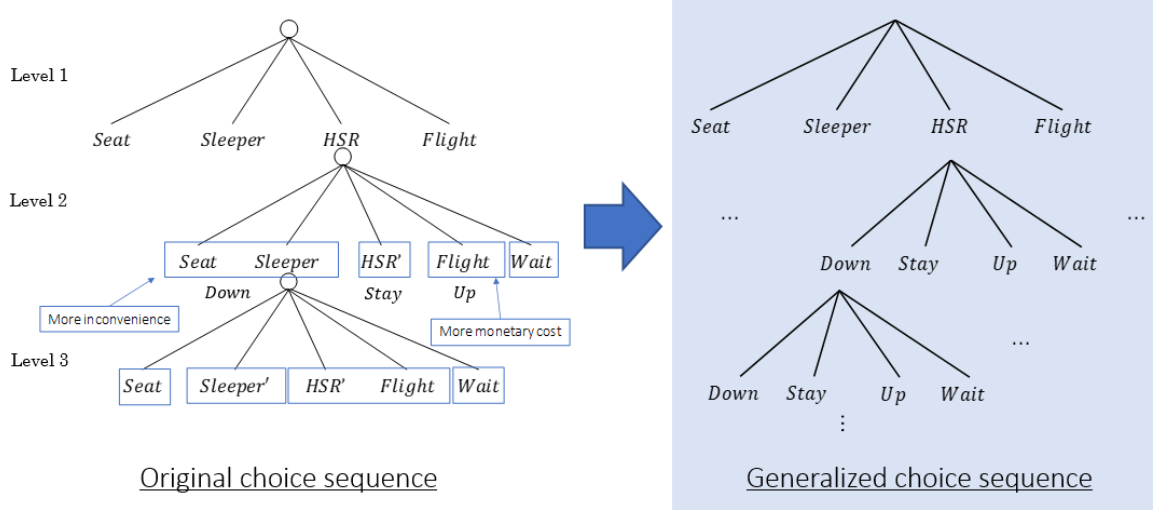


Figure 7: Illustration of the original and simplified sequential choice model

5.2 Level 2 choice model

The *Service Quality Preference* used for the first choice model is excluded in this model, since the Level 2 decision making instead emphasizes whether a person is giving up to obtain a ticket on this day or whether he is willing to settle on less preferred options. We therefore instead define *Eagerness to go home* as latent variable for this choice. *Eagerness to go home* is measured by two indicators, based on two questions about eagerness to check for refund tickets sold later if one could not get a ticket at all and if one is only able to obtain a ticket that is not one's preferred choice. Both questions are posed on a 5-point scale with endpoints "Very reluctant to check for refund tickets" and "Very willing to check for refund tickets".

To consider the correlation in choices made between Level 1 and Level 2 further a variable named *choice consistency* is defined. This variable is obtained from the output of the first model, by adding up the probabilities of the unified modes in Level 2 according to the definitions of the alternatives. That is, let k denote the choice in Level 1. Then

$$x_n^{L2}(i) = \sum_{j \in G_i(k)} x_n^{L1}(j) \quad i = \{Up, Down, Stay\} \quad (6)$$

$$0 \quad i = Wait$$

Where $x_n^{L2}(i)$ is a measure of the likelihood of person n choosing option i in the second

level which is defined according to choice set $G_i(k)$ which includes the Level 1 options that are included in “Up”, “Down” or “Stay” relative to choice k . For example, the *choice consistency* of “Up” for persons who chose Sleeper is obtained by $x_n^{L2}(Up) = P_n^{L1}(HSR) + P_n^{L1}(Flight)$ since HSR and Flight are both “Up” options relative to the persons preferred choice. *Choice consistency* is found to have a significant positive effect towards the corresponding choice. In other words, respondents tend to stick to their preference found in the first choice.

Given correlation and model tests, to control and to explore the influences inherited from the first model, *HSR dummy* and *Flight dummy* are introduced. The resulting choice model is shown Figure 8. The latent variable *Eagerness to go home* is represented as oval, and observed variables are represented as rectangles. Solid arrows represent structural relationships and dashed arrows represent measurement relationships. In the same way as for the Level 1 model, we model the indicators by ordinal logit and structure them by a linear formulation. The estimation results are presented in Table 5.

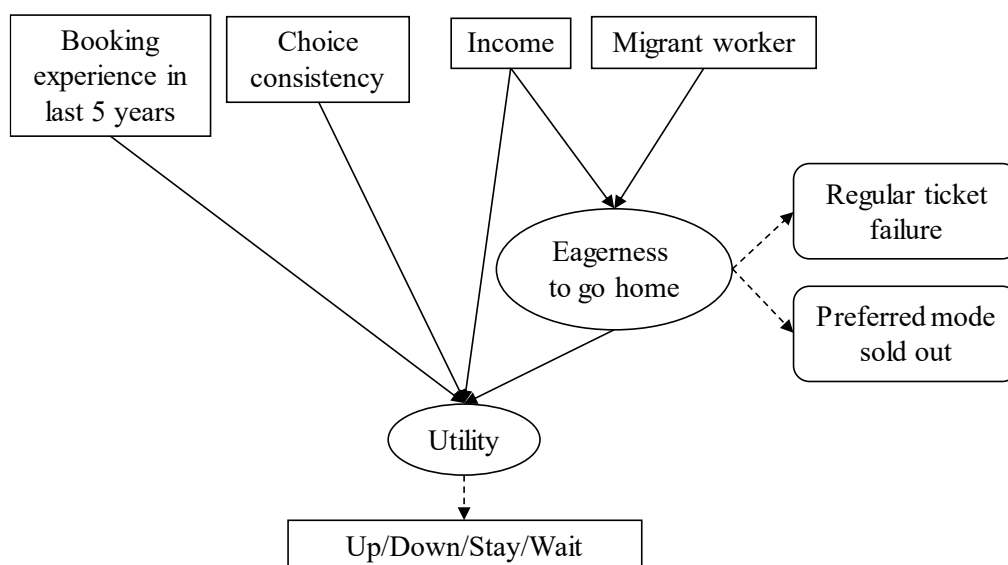


Figure 8: Framework of the Latent MNL model for alternative choice if preferred choice is not available (Level 2 choice)

Table 5: Parameter estimation results of Level 2 model

	Wait	Up	Stay	Down
	Estimation t-test	Estimation t-test	Estimation t-test	Estimation t-test
Alternative specific constant	ref. /	1.34 (2.38**)	0.713 (1.22)	2 (3.22***)
Middle income	ref. /	0.336 (0.669)	0.423 (0.86)	0.682 (1.34)
High income	ref. /	2.32 (2.73***)	1.75 (2.06**)	1.54 (1.74**)
Experienced in 5 years	ref. /	-1.34 (-2.46***)	-1.03 (-1.94*)	-1.1 (-2.01**)
Consistency	ref. /	2.47 (7.49***)		
Flight as 1 st choice	ref. /	ref. /	2.28 (4.02***)	ref. /
HSR as 1 st choice	ref. /	ref. /	ref. /	-1.05 (-3.04***)
Latent Variables				
Eagerness to go	ref. /	0.424 (1.19)	0.789 (1.98**)	0.773 (1.98**)
Structural Equations for Latent Variables				
	Estimation	t-test		
Middle income	-0.468	-1.9*		
High income	-0.473	-1.72*		
Migrant Worker	1.1	1.84*		
σ	0.761	3.49***		
Measurement Equations for Latent Variables				
	Estimation	t-test		
Regular ticket failure	intercept	4.46 9.78***		
	λ	1 /		
	τ_l	1 /		
	δ_l	2.16 7.95***		
	δ_2	1.64 9.31***		
	δ_3	2.27 10.7***		
Preferred mode sold out	intercept	4.69 7.15***		
	λ	1.79 1.99**		
	τ_l	1 /		
	δ_l	2.11 6.69***		
	δ_2	2.25 7.23***		
	δ_3	2.8 7.31***		
Model Statistics				
Sample Size	452	Likelihood ratio test		2480.99
Initial log likelihood	-2983.704	Rho-square		0.416
Final log likelihood	-1743.208	Adjusted Rho-square		0.405

* Significant at 10% ; ** Significant at 5% ; *** Significant at 1%

5.3 Result discussion

With *Eagerness to go home*, one can quantify that in general the more a person desires returning home, the less probable s/he chooses “Wait”. We obtain accordingly positive signs for all three options compared to Wait. The estimate is larger for “Stay” and “Down”, meaning that those who are more eager to go home will also more likely consider less convenient options.

The estimate for “Up” is not statistically significant. Looking at the measurement equations, we firstly note that the thresholds are statically significant, supporting our latent construct. Furthermore, eagerness to go home has more impact on “prefer mode sold out”, compared to “regular ticket failure”. In other words, those with higher eagerness to go home will not only check for alternative ways to go home if they fail to obtain any ticket but also if they did not obtain their preferred mode.

For the structural equations of the latent construct, we find that both middle and high income have a negative effect on eagerness to go home, this may be because people with higher income are more likely to be settled at the place they are living, hence having their own social circle and therefore less desire to go home. Even controlling for income, migrant workers show a strong desire to return home also in this choice level. Combined with results from the Level 1 model, migrant workers are found to be significantly more motivated to obtain a ticket compared to others. This is reasonable considering their working places are often thousands of kilometers from home and usually the Spring Festival vacation is the only chance for them to return every year.

Returning to the main structural model that determines the choice utilities, *Income dummy* is no longer included as it was not found to have a significant influence so that *Income* becomes the only monetary factor. With “Wait” as reference, *High income* has a significantly positive sign for other choices, meaning that those with higher income are more likely to switch to any reachable alternatives than giving up. Furthermore, considering the standardized coefficient, the positive impact on “Up” is greatest, followed by “Stay” and then “Down”. We also found that people who had a booking experience in the last five years tend to give up instead of attempting any further. An explanation might be that these persons have a lower expectation of the chance to obtain tickets based on their experience. Alternatively, one might interpret this as those who have returned home in some of the previous years feel less need to return home this spring festival as well. We acknowledge that both explanations remain tentative and require further tests.

Estimation results for the *HSR dummy* coefficient show that people who choose HSR are less likely to choose “Down”, i.e. are relatively reluctant to choose a non-HSR train. Also, the *Flight dummy* suggests that people who prefer to fly tend to choose Stay. Both observations confirm the discussion in Section 3.3 on dominant choice sequences. Finally, we note that the model fit is lower than for the Level 1 model though we suggest still acceptable. We conclude that it appears more difficult to predict people’s choice if the first preference is sold even in this significantly simplified choice scenario considering to the whole range of available alternatives.

6. Applying the choice models in an explorative simulation

Based on the above choice model results, simulations are conducted to investigate impacts of alternative policies as to ticket contingents and their order of sale. Clearly these results will be only explorative given our limited survey sample as well as limited questions, but we suggest that some general trends can be observed.

6.1 Simulation Framework

The overall simulation framework is shown in Figure 9. To firstly generate a random population sample the household income distribution from the Chinese Household Income Project 2013 (CHIP 2013) and age distribution from the Sixth National Population Census of China (Nation Bureau of Statistic of People’s Republic of China, 2010) are used to obtain fairly realistic distributions of these variables. To reflect correlation between age and income we utilize a linear regression between the two based on our sample. To generate the age for a person with household income category n we then utilize these regression parameters considering how much of the age distribution is explained by income and for the unexplained part we add a random error term drawn from a uniform distribution between 18 and 70. Other sociodemographic factors required in the model to define agents are generated based on their correlation with both income and age with a similar approach using linear regression for continuous variables or logistic regression for binary variables such as “spring festival experience”, or “migrant worker”.

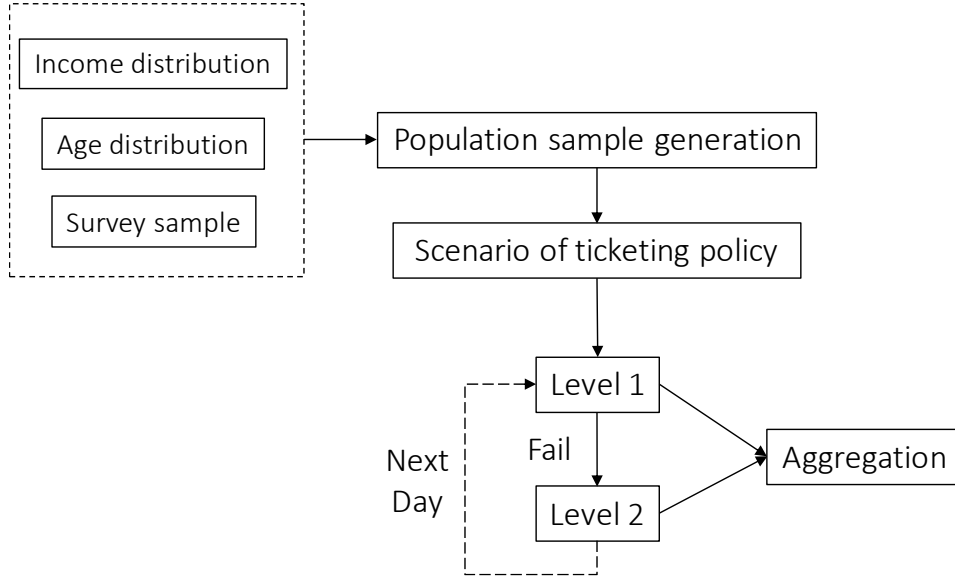


Figure 9: Flow chart of the simulation process

Input to the simulation are the parameters obtained from above models, a specific ticketing policy and a sample population. In all simulations conducted in this study, the population sample size N is fixed to 1000 and the ratio of Seat to Sleeper is fixed to 2:1, according to a typical layout of long-distance passenger trains in China (“China Railways rolling stock 25”).

By applying the framework, agents in the simulation will try up to two times to obtain their preferred and next preferred ticket before being forced to give up if they fail in both cases. We note that compared to the choice models in Section 4 two additional simplifications are made. Firstly, in order to control the capacity of the ticket pool, the “Stay” tickets which come from different ticket pools are excluded in the simulation as we only model one origin destination relation. Secondly, since the Level 2 choices are generalized, some of these choices become ambiguous. For example, a person with preference for *Seat* and *Up* could end up with a Sleeper, HSR or even Flight ticket. Supported by the analysis of dominant sequences shown in Figure 4, we presume that an individual chooses the next expensive option, i.e. *Seat* and *Up* means a person in our simulation will aim for *Sleeper* in Level 2.

6.2 HSR expansion scenario

With the rapid development and expansion of China’s high-speed railway network, an increasing number of “standard” railway routes are replaced with such faster but also more

expensive routes. This has led to public discussions that there are people being “priced-out” on longer journeys. Following this argument and our choice model results, a hypothesis is tested that if HSR replaces standard railway to a certain extent, a negative impact may appear for poor travelers and possibly the total number of people being able to obtain tickets reduces, hence, overall also leading to a decline in total number of successes.

The simulation is conducted under preset conditions as follows. First, as a specific constraint for this scenario, the sum of the total railway capacity (includes HSR, Sleeper and Seat) is fixed to 600 with a population of 1000. We vary the HSR capacity and adjust the Seat and Sleeper capacity accordingly (with fixed Seat to Sleeper ratio of 2:1 as noted before). Third, the capacity of Flight is set to be unlimited.

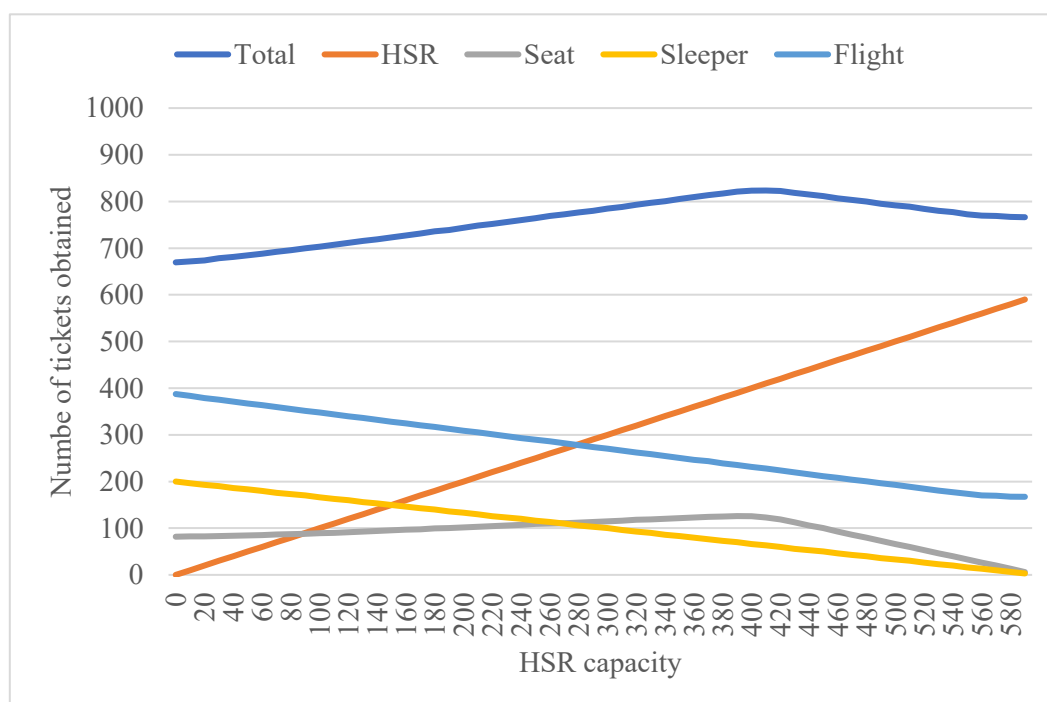


Figure 10: Tickets obtained depending on HSR capacity.

Results of the simulations are presented in Figure 10. The horizontal axis denotes the capacity of HSR in the simulation environment, which increases from zero (only standard trains) until all trains are converted to HSR (n standard trains). The vertical axis denotes the number of persons successfully obtaining a ticket. The lines are obtained from 60 data points, each of which is constructed from 900 simulation runs (30 random population draws x 30 choice draws

according to the logit model).

The straight lines for Sleeper and HSR show that these options are always sold in any scenario. The results show further that there are extreme points for the number of successes. At low HSR capacity the number of successes increases with increasing HSR capacity. At an HSR capacity of 400 the number of successes begins to drop because of those being financially constraint to take Sleeper or Seat options not being able to buy tickets anymore.

The results illustrate that replacing standard rail with HSR is beneficial only up to a limit. We show that too few standard rail capacity has a negative impact on the number of persons successfully obtaining a ticket as low income groups have less chance to return home during the Spring Festival travel season. We emphasize that we obtain the latter result, despite setting flight capacity to unlimited.

6.3 Staggered ticket sales scenario

Following our observation regarding low-income people, and to explore if ticketing policy can be used to increase the chance of this population group to obtain seat tickets, a second set of simulations is conducted. We aim to test if encouraging some of those persons who are fairly indifferent to several choice options to take their (slightly) less preferred tickets by modifying the order of ticket sales. By this some space might be extruded for the low-income population segment who can only afford seat tickets.

Table 6: Staggered ticket sales policies

	Scenario 1 Same time		Scenario 2 HSR first		Scenario 3 Sleeper & Seat first	
Ticket pool	Day1	Day2	Day1	Day2	Day1	Day2
HSR	400	Leftover from day1	400	Leftover from day1	0	400
Seat	133	Leftover from day1	0	133	133	Leftover from day1
Sleeper	66	Leftover from day1	0	66	66	Leftover from day1
Flight	Unlimited					

We therefore simulate staggered ticket sales. The ticket pool variables are set to vary by day as in the three scenarios presented in Table 6. Scenario 1 simulates the current situation, in which all the railway tickets for the same departure day are released at the same time. Scenario 2 is a policy to sell HSR tickets on Day 1 and normal rail tickets on Day 2. Vice versa in Scenario 3. All other preset conditions are the same as in the first simulation.

The results are presented in Table 7. In the HSR first policy, some of the people who would choose Seat or Sleeper as their first choice book HSR tickets in the first day, consequently the standard rail seat tickets are far from sold out in the simulation. For a similar reason, in the Sleeper & Seat first policy, HSR tickets remain available.

We clearly acknowledge that the numerous assumptions made in this simulation do not mean that we claim Seat or HSR tickets will actually be unsold. (For one, also some people might aim to return their “back-up ticket” and obtain their favorite ticket once it becomes available). However, what we suggest is that the simulation shows that the staggered sales can at least to some degree help reducing the pressure on the most demanded tickets.

Table 7: Number of success and total social utility under ticketing scenarios

	Number of Success				
	Seat	Sleeper	HSR	Flight	Total
Total ticket pool	133	66	400	Unlimited	/
Same time	132.9	66	400	296.4	895.3
HSR first	53.0	66	400	346.5	865.6
Sleeper & Seat first	133	66	281.5	461.9	942.4

7. Conclusions and further work

We described in this paper the problem travelers in China frequently face during the Spring Festival travel season. We discussed the scale of the demand and with it the severe capacity shortage during that time in the Chinese long distance public transport system. Our focus has been the initial preference and the change in preference of travelers once they experience a failure in obtaining a ticket. To model this, we introduce a sequential choice model rather than an integrated two level, nested choice model. The implication of this is that decisions are not “strategic”. The first choice is done irrespective of the chance of failure. We suggest this is a reasonable assumption given that there are few “investment costs” in trying to obtain one’s preferred option. That is, firstly, if one fails to obtain a ticket in the first place there is no monetary cost incurred and, secondly, there are a number of rebooking options independent of one’s first choice. Our two choice models are instead linked by including “choice consistency”, which we observe to be highly significant indicating that travelers will aim to obtain similar tickets in their second trial.

We suggest that the problem provides a framework for ticket choice under severe capacity shortage on which we found in general relatively few literature. The framework described in this study might be applicable also for other contexts where individuals have to make the trade-off between purchasing an inferior alternative or waiting for another chance to obtain their preferred option.

Looking more specifically into the observed preferences for our problem, we find a general preference for HSR tickets. This is not surprising given its “balanced” attractiveness in terms of speed (compared to standard rail) and price (compared to air). was not surprising. We found also that there appear to be separate markets for standard rail, sleeper and flight tickets. We showed with our “Level 1” choice model that a latent variable describing the preference for high service quality can describe the preferences well even controlling for income effects.

To reflect the 2nd choice we observe that “eagerness to go home” can explain the willingness people have to either spend more money or to travel longer/with less comfort. We could demonstrate that certain population groups, such as migrant workers, are more willing to choose less preferred options before “giving up”. We then utilized these findings to create a simulation with a sample population roughly reflecting Chinese socio-demographics. The simulation led to a number of insights. We showed that replacing standard rail with the faster and more comfortable, but also more expensive, HSR does not necessarily increase the number of people obtaining tickets. In addition we showed that staggered ticket sales might increase the total number of people being able to obtain tickets.

Clearly our work has a number of limitations that lead to further work directions. For one, our choice scenario was limited to a specific origin destination relation so that we could not distinguish time and cost trade-offs from the alternative specific constants. We also acknowledge that a number of simplifying assumptions were made as discussed throughout the manuscript. We note that factors such as the rise of “ticket brokers” whom one can pay for efforts to obtain desired tickets will have changed the problem for some. Further, the 2-level choice model might be extended to more days, though larger sample sizes would be desirable for this. Given our findings, if the survey would be repeated, we suggest respondents might be asked directly for their choice set and then to define their preferences and strategies in a nontrivial way within this choice set if certain options are sold out.

Acknowledgments

We acknowledge the valuable comments of the reviewers that significantly helped improving this manuscript. Different parts of this paper have been presented at the 7th INSTR (International Symposium on Transportation Network Reliability) and at the 15th IATBR (International Conference on Transport Behavior Research). We thank the reviewers of both abstracts and the audiences during the presentations for their comments.

Conflict of Interest

Both authors confirm that there is no conflict in interest.

Author contribution

Yuhan Gao: Data collection and analysis, manuscript writing
Jan-Dirk Schmöcker: Content design, manuscript writing and editing

References

- Bahamonde-Birke, F. J., & Hanappi, T. (2016). The potential of electromobility in Austria: Evidence from hybrid choice models under the presence of unreported information. *Transportation Research Part A: Policy and Practice*, 83, 30-41.
- Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.
- Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. *Perpetual motion: Travel behaviour research opportunities and application challenges*, 431-470.
- Bierlaire, M. (2018). Estimating choice models with latent variables with PndasBiogeme (No. REP_WORK).
- Bierlaire, M. (2018). PndasBiogeme: a short introduction (No. REP_WORK). Technical Report TRANSP-OR 181219, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.
- Chen, T. S., Mao, B, Gao, L. and Yue, F. (2007). Research about Passenger Travel Choice Behavior of Dedicated Passenger Railway Line. *Journal of the China Railway Society*,

29(3):8-12.

- Cheng, Y. H. and Huang, T. Y. (2014). High speed rail passenger segmentation and ticketing channel preference. *Transportation Research Part A: Policy and Practice*, 66: 127-143.
- China Household Finance Survey (2011). Available from Survey and Research Center for China Household Finance, <https://chfs.swufe.edu.cn> [Accessed Aug 2017].
- Chinese Household Income Project (2013). Available from China Institute for Income Distribution, <http://www.ciidbun.org> [Accessed Aug 2017].
- Chorus, C. G., Arentze, T.A. and Timmermans, H. J. (2008). A random regret-minimization model of travel choice. *Transportation Research Part B: Methodological* 42.1 (2008): 1-18.
- Chorus, C. G., and Kroesen, M. (2014). On the (im-) possibility of deriving transport policy implications from hybrid choice models. *Transport Policy*, 36, 217-222.
- Daly, A., Hess, S., Patruni, B., Potoglou, D., & Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation*, 39(2), 267-297.
- Dellaert, B. G., Borgers, A. W., and Timmermans, H. J. (1997). Conjoint models of tourist portfolio choice: Theory and illustration. *Leisure Sciences*, 19(1): 31-58.
- Eckstein, Z. and Wolpin. K. I. (1989). The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, 24(4): 562-598.
- Eymann A, Ronning G and Zimmermann K F. (1992). Discrete choice analysis of foreign travel demand. *European integration in the world economy*. Springer, Berlin, Heidelberg, 641-674.
- Fosgerau, M., Frejinger, E. and Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56: 70-80.
- Fu, H. and Chester, W. (2004). Sequential logit dynamic travel demand model for hurricane evacuation. *Transportation Research Record: Journal of the Transportation Research Board*, 1882: 19-26.
- Liu, Y., and Cirillo, C. (2017). A Generalized Dynamic Discrete Choice Model for Green Vehicle Adoption. *Transportation Research Procedia*, 23: 868-886.
- Oyama Y. and Hato E. (2017). A discounted recursive logit model for dynamic gridlock network

- analysis. *Transportation Research Part C: Emerging Technologies*, 85: 509-527.
- Quiggin, J. (1994). Regret theory with general choice sets. *Journal of Risk and Uncertainty* 8(2): 153-165.
- Sanko, N., Hess, S., Dumont, J., & Daly, A. (2014). Contrasting imputation with a latent variable approach to dealing with missing income in choice models. *Journal of choice modelling*, 12, 47-57.
- Scales, J., Ollivier, G. and Amos, P. (2011). Railway price regulation in China: time for a rethink? *China Transport Topics* No.01.
- Shao, C. (2008). A discussion on migrant worker flow in Spring Festival. *Journal of Anhui Agricultural University (Social science edition)*, 17(6): 57-61.
- State Council of the People's Republic of China. (2017). 2017 年春运客运量逾 29 亿再创新高. (In Chinese)
- Timmermans, H. J. (1996). A stated choice model of sequential mode and destination choice behaviour for shopping trips. *Environment and Planning A*, 28(1): 173-184.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Urata, J., and Hato, E. (2013). An evacuation-chain based dynamic assignment model with evacuation timing and delay penalty. *Proceedings of the 18th International Conference of Hong Kong Society for Transportation Studies (HKSTS 2013)*: 573-578.
- Wang, J. F., & Huang, R. H. (2016). Can You Get a Ticket? Adaptive Railway Booking Strategies by Customer Value. *Journal of Public Transportation*, 19(4): 1-17.
- Zhou, Y. (2013). 春运火车票应如何定价?——从法与经济学角度分析. *中国证券期货*, (6X): 210-210. (In Chinese).
- Zhou, Y. and Kang, Y.. (2006). Economic Analysis and Systematic Innovation of Train Ticket Selling during Spring Festival. *Journal of Hunan University of Arts and Science (Social Science Edition)*, 31(3): 83-87.