Modelling the cancellation behaviour of hotel guests

Falk, Martin; Vieru, Markku

Published in:
INTERNATIONAL JOURNAL OF CONTEMPORARY HOSPITALITY MANAGEMENT

DOI:
10.1108/IJCHM-08-2017-0509

Published: 08.10.2018

Document Version
Peer reviewed version

Citation for pulished version (APA):
Falk, M., & Vieru, M. (2018). Modelling the cancellation behaviour of hotel guests. *INTERNATIONAL JOURNAL OF CONTEMPORARY HOSPITALITY MANAGEMENT*, *30*(10), 3100-3116. https://doi.org/10.1108/IJCHM-08-2017-0509

Modelling the Cancellation Behaviour of Hotel Guests

Martin Falk

Austrian Institute of Economic Research (WIFO)
Arsenal Objekt 20
A-1030 Vienna, Austria
Tel. +43 664 4490504
E-mail: martin.falk@wifo.ac.at

Markku Vieru
University of Lapland, Multidimensional Tourism Institute (MTI)
P.O. Box 122,
FI-96101 Rovaniemi, Finland
Tel. +358 400 37641
E-mail: markku.vieru@ulapland.fi

Abstract:

Purpose – The aim of this study is to provide new insights into the factors that influence cancellation behaviour with respect to hotel bookings. The data is based on individual bookings drawn from a hotel reservation system database comprising nine hotels.

Approach – The determinants of cancellation probability are estimated using a probit model with cluster adjusted standard errors at the hotel level. Separate estimates are provided for rooms booked offline, through online travel agencies, and through traditional travel agencies.

Findings – Evidence based on 233,000 bookings shows that the overall cancellation rate is 8 per cent. Cancellation rates are highest for online bookings (17 per cent), followed by offline bookings (12 per cent) and travel agency bookings (4 per cent). Probit estimations show that the probability of cancelling a booking is significantly higher for early bookings, large groups that book offline, offline bookings during high seasons, bookings not involving children, and bookings made by guests from specific countries (e.g. China and Russia). Among the factors, booking lead time and country of residence play the largest role, particularly for online bookings.

Research limitations/implications – The analysis is based on individual-level booking data from one hotel chain in Finland, and therefore cannot be generalised for the total population of hotels in the country under observation.

Originality/value – The main contribution of this paper is a thorough investigation of the factors that influence cancellation behaviour at both the theoretical and empirical level. Detailed and unique data from a hotel reservation system allows for new empirical insights into this behaviour.

Keywords: hotel management, cancellations, booking channel, online booking, probit model

Acknowledgments: The authors would like to thank the editor, the anonymous reviewers, Peter Björk, Kaye Chon, Daniel Fesenmaier, Eva Hagsten, Raija Komppula, Cristian Morosan, Fevzi Okumus, Serena Volo, and Malin Zillinger and the participants of the CBTS2016 conference in Brunico, the Future of Winter Tourism conference (FWT2017) in Rovaniemi, the TTRA2017 conference in Angers, and the 26th Nordic Symposium of Tourism and Hospitality Research 2017 in Falun for their helpful comments on an earlier draft of the paper. We also thank the hotel for providing the booking data. Tyler Schaffner and Tess Landon are acknowledged for excellent proofreading.

1 Introduction

A significant number of hotel bookings are not realised due to cancellations or no-shows. The typical reasons include sudden illnesses, accidents, schedule conflicts, unexpected (family) obligations, and natural catastrophes. Although the option to cancel a hotel booking (preferably at a low cost) is beneficial to presumptive hotel guests, it is a less desirable and possibly revenue-diminishing factor for hotel managers to deal with (Chen and Xie, 2013). Such losses are particularly high on last-minute cancellations (Chen, Schwartz, and Vargas, 2011; Koide and Ishii, 2005). However, new technologies involving online booking channels have dramatically changed customers' booking possibilities and behaviour (Dolnicar and Laesser, 2007; Lee, Guillet, and Law, 2013). This adds a further dimension to the challenge of how hotels handle cancellations, which are no longer limited to traditional booking and guest characteristics.

This study offers both a theoretical framework and a quantitative analysis of factors that influence cancellations of booked hotel rooms. Both guest- and booking-specific characteristics are investigated, and the estimations are performed using a probit model. The unique dataset at hand originates from a hotel reservation system comprising nine hotels and approximately 233,000 bookings over a five-year period. Eight per cent of the bookings are recorded as cancelled in the system.

In general, little is known about the cancellation behaviour of travellers and hotel guests (Hajibaba, Boztuğ, and Dolnicar, 2016). Historically, the cancellation rate of hotel bookings has been lower than that of booked flights (DeKay, Yates, and Toh, 2004; Toh, 1986). However, the free cancellation option provided by online travel agents (OTAs) may lead to higher cancellation rates. Anecdotal evidence suggests that cancellations of rooms booked online have increased recently due to "click to cancel" options (Delgado, 2016). Previous studies have often focused on the cancellation policies and booking behaviour of hotels (Chen et al., 2011; Chen and Xie, 2013); the relationship between cancellation fees, time of booking, and the intention to cancel (Park and Jang, 2014); and actions taken by destination managers to prevent tourists from cancelling (Hajibaba et al., 2016). Cancellation probability is also a key parameter in theoretical models of overbooking (Koide and Ishii, 2005; Subramanian, Stidham, and Lautenbacher, 1999).

Few studies have investigated the determinants of the probability of hotel cancellations using statistical models. Based on 240,000 booking records, Morales and Wang (2010) find that several guest-, booking- and room-specific characteristics are relevant in forecasting cancellation rates. Antonio, de Almeida, and Nunes (2017) show that booking channel, arrival month, room type, booking lead time, and country of origin are the most important predictors. Meanwhile, several studies have investigated the determinants of cancellations and no-shows for flights (Garrow and Koppelman, 2004; Iliescu, Garrow, and Parker; 2008; Xiong and Hansen, 2013). However, passenger name record (PNR) systems for air travellers are less detailed than hotel booking databases, which include each guest's itinerary, departure day and time, reservation method, booking class, gender, and name. Studies modelling no-show or cancellation behaviour are also available for restaurant visits (Tse and Poon, 2016).

Knowledge of cancellation behaviour is relevant not only to hotel managers' predictions of future revenues and capacity utilisation, but to their cancellation and pricing policies, as well. Obtaining such knowledge, however, places high demands on the data source in question, which needs to be rich and (preferably) also correspond to the flow format of a hotel booking system or related databases drawn from enterprise resource planning (ERP) systems. Booking data drawn from hotel booking systems can be an important pillar of destination management information systems that collect data from different sources (Höpken, Fuchs, and Lexhagen, 2014; Höpken et al., 2015). These studies use an explorative-inductive approach to explain consumers' cancellation behavior and employ a decision tree algorithm approach on destination wide booking data.

This study contributes to the literature by providing further results on the role various guest and booking characteristics play in cancellation probability. Specifically, after formulating several hypotheses we estimate the significance and magnitude of various cancellation determinants by each booking channel. This work is thus more detailed than previous studies by Morales and Wang (2010) and Antonio et al. (2017), who put more emphasis on forecasting booking cancellations.

This paper is structured as follows. Section 2 introduces the theoretical background and the hypotheses, while section 3 introduces the empirical model. Section 4 presents the data and descriptive statistics. Section 5 presents empirical results, and section 6 concludes.

2 Theoretical background and formulation of hypotheses

Several factors – whether connected to guests' decisions or motivations to travel or the booking process itself – can affect the cancellation of a hotel booking. In most cases, the travel decision (Sirakaya and Woodside, 2005) is made first and then potentially followed by a hotel reservation. In this context, a guest i) determines booking-specific details (such as the booking channel used, arrival date, length of planned stay, motivation to travel, and room category) and ii) provides guest-specific details (such as his or her name, credit card number, group size, number of children, and country of residence), which are then stored in the hotel's booking system.

It is inevitable that advance bookings include elements of uncertainty. The cancellation of a booking can be defined as an instance in which a presumptive hotel guest encounters a specific

event or incident that has a significant negative effect on his or her willingness to complete a planned trip (Chen et al., 2011; Chen and Xie, 2013; Chew and Jahari, 2014; Hajibaba et al., 2015). Such instances can involve, for example, changes in vacation plans, a sudden illness, unfavourable weather conditions, a change in employment, and rescheduled business meetings. Under ideal conditions that assume no uncertainty and risk-averse individuals who maximise rational utility, no cancellation exists. Deviation from these conditions leads people to re-evaluate their travel plans. Guests can either inform the hotel that they do not intend to travel prior to the arrival date (“cancellations”) or simply not check in (“no-shows”).
Prior studies show that some bookings are more likely to be cancelled than others. This may be due to differences in risk perceptions (Cheron and Ritchie, 1982; Moutinho, 1987; Hajibaba et al., 2016; Sharifpour et al., 2014), one’s level of information or familiarity with the destination in question (Lepp and Gibson, 2003), and the flexibility of travel plans (Fesenmaier and Jeng, 2000; Park and Fesenmaier, 2014). In their conceptual model, Hajibaba et al. (2015) introduce a segment of tourists who seldom cancel their bookings.

The literature has identified several factors that determine the probability of cancellation. First, the booking channels guests use may be associated with how likely they are to cancel. The choice of channel is typically not random, but related to individual or technical circumstances (Dolnicar and Laesser, 2007) such as travel budget, information needs, and processing time (Law, Leung, and Wong, 2004). According to Cheyne, Downes, and Legg (2006), online channels are often used to book simple, short-term, or budget-minded trips. Long-distance travellers, on the other hand, tend to favour travel agencies, since these trips are often characterised by a high degree of complexity. Castillo-Manzano and López-Valpuesta (2010) arrive at similar conclusions. The tailored service and destination-specific information provided by a travel agent is likely to reduce the risk of cancellation in comparison to the impersonal service provided by online booking sites.

Online booking has many advantages, but the online environment may also result in impulsive decisions and that lead to a higher rate of cancellation. For example, certain features of booking sites (e.g. attractive interactive graphics, product descriptions, and the ease of booking and cancelling) may weaken the self-control of certain individuals (LaRose and Eastin, 2002). In addition, Donthu and Garcia (1999) find that online shoppers are more impulsive than others. This is supported by a study on buying and selling stocks online (Barber and Odean, 2002), implying that easily accessible online channels may attract more unfulfilled bookings than do bookings made offline.

The online booking environment can modify the trip decision model. According to Fesenmaier and Jeng (2000) the core elements of holiday trips are planned ahead of time, while secondary trip-related decisions remain tentative and flexible. Guests are likely to be strongly committed (in relative terms) after booking via travel agencies due to the significant cancellation penalties involved. Bookings via OTAs, on the other hand, are relatively simple and cheap to modify without direct personal contact with hotel staff. In addition, new information guests obtain on room pricing after booking may influence their decisions to cancel (Chen et al., 2011). It is well known that price transparency is higher in the online booking environment (Kim, Cho, Kim, and Shin, 2014), which is another factor that suggests OTA bookings are more likely to be cancelled than offline reservations. Empirical evidence also bears out the importance of the distribution channel at hand in predicting cancellations (Morales and Wang, 2010; Antonio et al., 2017). These arguments lead to our first hypothesis, namely that the cancellation probability is the highest for rooms booked online (OTA) and the lowest for rooms booked through travel agencies.

Second, cancellation rates may vary based on booking lead time. Basically, cancellation can take place any time after a given booking. When there is a long time interval between the booking and the check-in date, circumstances affecting the booking are more likely to arise (Chew and Jahari, 2014). Bookings made far in advance are more exposed to unpredictable internal and external events that are not fully considered on the date of booking (Fesenmaier and Jeng, 2000; Hajibaba et al., 2015). Since the majority of booking channels charge cancellation fees after a certain date, cancellation rates tend to decrease drastically after this date passes (Zakhary et al., 2011). Prior empirical evidence further establishes booking lead time as a predictor of cancellations (Morales and Wang, 2010; Antonio et al., 2017). Consequently, our second hypothesis states that booking far in advance increases the probability of cancellation.

Furthermore, attitudes towards uncertainty and other cultural differences between countries (as described by Hofstede, 2001) may also explain why the risk of cancellation varies by country of residence. Law (2006) finds that Asian travellers are more likely to adjust their travel plans than are visitors from other countries. Similarly, Kozak, Crotts, and Law (2007) suggest that travellers from specific cultures (e.g. Singapore and China) have a lower risk tolerance and are more likely to change their travel plans than are visitors from cultures more accepting of risk (see also Seddighi, Nuttall, and Theocharous, 2001). Antonio et al. (2017) show that country of residence is a major factor in predicting cancellation rates. Our third hypothesis thus states that country of origin affects the probability of cancellation.

Cancellations may also correlate to the amount of travel distance at hand. Here, guests who travel to faraway destinations are presumably seeking novelty and excitement. Roehl and Fesenmaier (1992), for example, state that uncertainty can be part of the excitement of a trip rather than a problem. Lepp and Gibson (2003) also suggest that novelty-seekers tolerate a higher level of uncertainty than those who prefer familiar surroundings. Similarly, guests travelling within or near their home country tend to be better informed about their destination, pay less for the trip, view it as less exotic, are more likely to change their travel plans (Park and Fesenmaier, 2014), and are thus more liable to cancel their plans. Our fourth hypothesis thus states that guests travelling long distances are less likely to cancel bookings than are domestic guests or guests from neighbouring countries.

Seasonality is considered another important factor in determining cancellation behaviour, as occupancy varies between the high and low seasons. In the high season, the probability that a guest's preferred room will not be available or a given hotel will be fully booked is relatively high (Nicolau and Masiero, 2017). Therefore, guests tend to book hotel rooms even when they are not completely sure that they can make the trip. The logic behind this is that there will be enough time to think about the decision and cancel the booking if necessary. In addition, there is social pressure to go on holiday in the high season. According to decision models (e.g. Ajzen, 1991; Moutinho, 1987), in certain social environments, e.g. holiday and peak seasons, may involve relatively high pressure to act according to the perceived norms of a reference group, which leads to more bookings. Given these motivations, our fifth hypothesis posits that cancellation probability is higher during peak seasons.

Guest-specific factors (such as the composition of a given travel group) are also expected to influence cancellations. As pointed out by Sirakaya and Woodside (2005), tourism is a social activity that often involves family or friends. This implies that cancellation behaviour may depend on how cohesive a given group is (Bollen and Hoyle, 1990). More loosely established groups with less homogeneous attitudes, beliefs, and values are assumed to have a higher cancellation rate (Milliken and Martins, 1996). Park and Fesenmaier (2014) extend this discussion by pointing out that large groups also mean more individuals with different preferences, which could increase the risk of cancellation. On the other hand, families with children may cancel to a lesser extent. So and Lehto (2007) distinguish between travelling with family, with friends, or alone and suggest that groups of friends are more prone to changing their plans. This is in concordance with Thornton, Shaw, and Williams (1997), who find that guests with children are less likely to cancel. However, Antonio et al. (2017) find that guests with babies do not play a role in predicting in cancellation rates. Consequently, our sixth hypothesis states that while large groups of adults pose a higher cancellation risk than others, bookings that include children are less likely to be cancelled.

In summary, the hypotheses this paper explores can be formulated as follows:

H1:	Cancellation probability is the highest for rooms booked online (OTA) and the lowest for rooms booked through travel agencies
H2:	Booking far in advance increases the probability of cancellation
H3:	Country of origin affects the probability of cancellation
H4:	Guests travelling long distances are less likely to cancel bookings than are domestic guests or guests from neighbouring countries
H5:	Cancellation probability is higher during peak seasons
H6:	Large groups of adults pose a higher cancellation risk than others, but bookings that include children are less likely to be cancelled

Furthermore, the purpose of a trip (i.e. leisure or business) may affect a customer's cancellation behaviour. On the one hand, business travellers can be expected to present a higher cancellation risk because their planned business meetings may change suddenly (Liu, 2004). However, business clients are often attending conventions and seminars where their presence is essential. Failing to participate in negotiations with clients regarding financial deals or business development, for example, can result in delays or financial losses. This normally leads to lower cancellation rates. Since the relationship between the probability of cancellation and the type of guest is not clear-cut, no corresponding hypothesis is explored here; instead, these factors are used as control variables.

Planned length of stay, booking and arrival day of the week, arrival month and arrival year, and room category are additional factors that may influence cancellations. These variables are considered in studies that analyse the forecasting of cancellations (Morales and Wang, 2010; Antonio et al., 2017). A longer stay presumably requires more careful planning, for example, and the weekdays on which a guest books a room and is scheduled to arrive can also affect the likelihood of cancellation. Finally, hotels have rooms that vary in size and other characteristics, all of which are likely associated with the probability of cancellation. These factors are treated as control variables, meaning no hypotheses are derived regarding how they relate to cancellations.

3 Empirical model

Presumptive guests who have booked a hotel room face two choices: i) check in or ii) cancel their booking (or simply not show up). No-shows cannot be explicitly modelled in this study because the share of no-shows in total cancellations is negligible with about 2 per cent. These two alternatives may be estimated by a probability model in which the outcome represents the individual decision of whether or not to cancel, which is assumed to be an unobserved latent variable. Here, the cancellation probability is modelled using a probit model (the individual index *i* is suppressed for the sake of convenience):

$$Pr(Y = 1 \mid X) = \phi(X'\beta), \tag{1}$$

where ϕ is the cumulative distribution function (CDF) of the standard normal distribution and the cancellation probability, Y^*, is a function of the observable characteristics, X:

$$Y^* = X'\beta + \varepsilon. \tag{2}$$

The underlying observed variable Y is a binary variable that is defined as follows:

$$Y = \begin{cases} 1 & if \quad Y^* > 0 \\ 0 & if \quad Y^* \leq 0 \end{cases} \tag{3}$$

X is a vector of covariates containing booking- and guest-specific characteristics, and β is the corresponding coefficient vector. Random factors, along with unobservable factors that influence cancellation decisions, are captured by the error term, ε.

Given the theoretical considerations based on earlier literature, for a given booking channel the cancellation decision Y is specified as a function of several factors:

$$\begin{aligned} Y_{it} = {} & \beta_0 + \sum_{BT=1}^{5} \beta_{BT} BOOKTIME_{ijtBT} + \sum_{S=1}^{4} \beta_S SEASON_{ijtS} + \sum_{A=1}^{3} \beta_A NOADULTS_{ijtA} \\ & + \beta_C CHILDREN_{ijt} + \sum_{CR=1}^{19} \beta_{CR} COUNTRY_{ijtCR} + \sum_{L=1}^{4} \beta_L LENGTHSTAY_{ijtL} + \sum_{BS=1}^{2} \beta_{BS} BUSINESS_{ijtBS} \\ & + \sum_{H=1}^{8} \beta_H HOTEL_{ijtA} + \sum_{CA=1}^{17} \beta_{CA} CATEGORY_{ijtCA} + \sum_{AM=1}^{11} \alpha_{AM} ARRIVALMONTH_{ijtAM} \\ & + \sum_{AJ=1}^{5} \alpha_{AJ} ARRIVALYR_{ijtAJ} + \sum_{DA=1}^{6} \alpha_{DA} DOWARRIVAL_{ijtDA} + \sum_{DB=1}^{6} \alpha_{DB} DOWBOOKING_{ijtDB} + \varepsilon_{it} \end{aligned} \tag{4}$$

where i denotes the individual bookings on arrival day t at hotel j. *BOOKTIME* is a set of dummy variables consisting of several categories that measure the number of days between booking and check-in dates (where *BOOKTIME1t4,* which denotes guests who made their reservation between one and four days before their arrival date, represents the reference category). Bookings made on the day of arrival are excluded. *SEASON* denotes a set of dummy variables for Christmas and New Year's, winter breaks, and Easter and summer school breaks, with the non-holiday season as the reference period. Christmas, New Year's, Easter, and winter break are the peak periods of tourism demand in the snow season. *NOADULTS* is a set of dummy variables that indicates the number of adult guests, with single bookers serving as the reference category. The dummy variable *CHILDREN* includes bookings of guests with children. *COUNTRY* is a set of dummy variables indicating country of residence, with Finland as the reference category. Furthermore, *LENGTHSTAY* denotes four dummy variables indicating the planned length of stay, with one-night stays representing the reference category. The specification also takes into account business travellers (*BUSINESS*), which are measured as two dummy variables indicating whether or not guests are business customers or members of an association. Here, individual bookers serve as the reference category (information is only available for offline bookings). *ARRIVALMONTH* measures monthly effects and *ARRIVALYR* denotes the arrival year, with both defined as sets of dummy variables. The early dummy variables contained in *ARRIVALYR* control for aggregate factors that vary over time, such as the business cycle and the inflation rate. Hotel dummy variables (*HOTEL*) capture time-invariant, hotel-specific factors such as location and quality segment. *CATEGORY* is a set of dummy variables indicating the size and quality of the hotel room in question based on 17 categories and the reference category. *DOWARRIVAL* denotes a set of dummy variables indicating the day of the week on which each guest checked in (Monday to Saturday), with Sunday as the reference category. Finally, *DOWBOOKING* denotes a vector of dummy variables that capture the weekday on which each guest booked, with Sunday again serving as the reference category.

The cancellation probability is estimated by the probit model using the maximum likelihood estimator (or, alternatively, by the logit model; Wooldridge, 2010). Standard errors are clustered at the hotel level to control for possible correlations across bookings in a single hotel (Wooldridge, 2010). Results are provided for the total sample over all booking channels as well as separate estimation results for each channel (offline, travel agencies, and OTAs) are provided.

4 Data and descriptive statistics

The data consists of individual bookings drawn from the hotel booking system of nine hotels belonging to a hotel chain located in Finnish Lapland for the period January 2011 to March 2016. It contains detailed information on guest- and booking-specific characteristics, including room prices, booking and arrival dates, dates of departure, room categories, number of guests (distinguishing between adults and children), and countries of residence (Falk and Vieru, in press). In addition, the database reveals which of the following booking channels was used to book a given room: i) online through the hotel's own platform, ii) online through OTAs, iii) offline using traditional travel agencies, or iv) direct contact with the hotel (phone, e-mail, front desk). Bookings made via travel agencies are not recorded in a timely manner. Therefore, the relationship between booking lead time and the cancellation rate should be interpreted with caution for this channel. Meanwhile, cancelling a booking can either be free or subject to a fee, which varies by booking lead time and booking channel. Information on guest type (business guests, members of associations, and non-business guests) is only available for offline bookings.

The total number of bookings in the database in question is about 300,000. Rooms booked via a hotel's website are excluded because corresponding cancellations are not registered in the booking system. Excluding these bookings reduces the sample size by 67,000 (to 233,000). Of the total number of bookings (including cancellations), 12 per cent are OTA bookings, and 17 per cent of these were not realised. Offline bookings account for 38 per cent of the bookings, with a cancellation rate of 12 per cent. Travel agencies are responsible for 51 per cent of the bookings and exhibit the lowest cancellation rate (4 per cent). The cancellation rates for online and offline bookings are consistent with Antonio et al. (2017), who report rates ranging between

12 and 26 per cent based on 73,000 bookings from four hotels. Morales and Wang (2010) find a cancellation rate of 20 per cent based on 240,000 booking records from one hotel in the UK.

Descriptive statistics show that the majority of guests arrived alone or with just one other person (Table 1). In addition, short stays are much more common than extended arrangements, and few guests book very early (defined as 100 or more days before the planned arrival date). The share of Finnish residents is 79 per cent for offline bookings and 55 per cent for bookings through OTAs. Guests from Germany account for the largest share of bookings arranged via travel agencies (34 per cent).

[Table 1 about here]

The descriptive statistics in Table 1 also reveal that cancellation rates vary widely by characteristics. In particular, cancellation rates increase with the length of stay and are highest in the winter holiday season. Guests who book very early exhibit the highest cancellation rate. For instance, guests who book 100 days or more before their arrival date cancel between 13 and 40 per cent of the time depending on the booking channel at hand. In contrast, guests who book between one and four days before arriving cancel at rates ranging between 3 and 9 per cent (again, depending on the booking channel).

5 Empirical results

The probit estimations in Table 2 show that the factors affecting the probability to cancel a hotel booking differ across channels. Among the factors identified, lead time and country of residence are the strongest determinants of cancellation probability. The share of correctly classified cancellations is about 92 per cent indicating a good fit. The likelihood of cancellation increases monotonically with booking lead time, meaning that late bookings are associated with a lower risk of cancellation. Overall, this is consistent with the theoretical prediction formulated in hypothesis two (H2). An exception can be observed in bookings made via travel agencies, which follow a U-shape pattern. Here, the cancellation probability is lowest for bookings made between 5-9 or 10-24 days prior to arrival. Early reservation increases the probability of cancellation for both online and offline bookings (by 52 and 11 percentage points, respectively, compared to the reference category of 1-4 days).

[Table 2 about here]

A higher cancellation probability is expected for early bookings because plans with long time horizons are naturally more likely to be changed due to unforeseen events. Possible reasons why customers cancel early bookings (which is often made easier by the option to do so free of charge) include changing plans, scheduling conflicts, and sudden illnesses. The finding that the booking lead time effect is much more pronounced for OTAs than for offline or travel agency bookings is consistent with related theoretical predictions (Donthu and Garcia, 1999). An F-test confirms that the coefficients of booking lead time are significantly higher for online than for offline bookings at the five per cent level.

The second most important factor concerning cancellation probability in Table 2 is the country of residence of guests. The variation between OTA and offline booking is about 30 and 20 percentage points, respectively (denoting the difference between the countries with the lowest and highest cancellation probabilities). Hypothesis H3 can thus be confirmed. This shows that bookers from different countries make different decisions with respect to changes in their travel plans, which confirms the general literature on differences in travel behaviour across countries (e.g. Law, 2006). In particular, travellers from the Netherlands, the UK, and Switzerland exhibit a significantly lower cancellation probability, whereas those from Russia and China are much more likely to cancel compared to the reference group (domestic travellers). However, long-distance travellers from Japan and other non-European countries (except Hong Kong, Singapore, and Taiwan) are not associated with a lower likelihood of cancellation, which leads us to reject hypothesis H4. The higher cancellation probability of Chinese and Russian guests who book online can be explained by the fact that a hotel booking confirmation is required to obtain a Schengen visa for the EU countries. These guests do not want to specify exactly in advance where they stay and therefore use more often the free to cancellation option. For travel agency bookings, the variation in cancellation probability by country of residence is significantly less pronounced.

Furthermore, offline bookers are more likely to cancel in peak seasons (Christmas week, New Year's, Easter holiday week) than in quieter periods, which is consistent with hypothesis H5. The differences correspond to about two percentage points, which is quite low compared to the other booking- and guest-specific characteristics. Since our model includes monthly dummy variables that partially account for holiday seasons, this result is not that surprising. The month of arrival is another significant factor that determines the likelihood of cancellation. For offline bookers, the cancellation probability in July and August is four percentage points lower than for the reference arrival month of January.

The relationship between cancellation rates and the number of guests is not clear-cut. For offline guests and travel agencies, estimates show that the larger the number of guests is in a given room, the higher the risk of cancellation. This is in line with the group cohesion hypothesis stated in H6. However, there is no clear relationship between the number of guests and the cancellation probability of online bookers.

Along with country of residence and booking lead time, length of stay plays an important role. For offline and travel agency bookings, our results show that guests who only plan to stay one night are much more likely to cancel (by three percentage points on average) compared to those who book two nights or more. As expected, OTA bookers who plan to stay more than three nights exhibit contrasting behaviour in the form of a cancellation probability that is four percentage points higher. This indicates that guests who have booked a longer stay online tend to be overly optimistic about their plans. The ease of booking online (LaRose and Eastin, 2002) seems to have led to a rise in impulsive bookings that are not based on carefully considered travel plans.

Meanwhile, presumptive business clients and association members who book offline have a significantly lower probability of cancellation (see Table 3 in appendix). The difference is large – about five percentage points compared to non-business guests. This is likely due to the opportunity costs of cancellation being higher for business clients than for non-business clients. In addition, the weekday of arrival – measured as a set of dummy variables that takes Sunday as its reference day – is jointly significant at five per cent level. However, marginal effects are quite low in the majority of cases. One exception can be seen in online bookers who are scheduled to arrive on a Thursday, who are 2.8 percentage points more likely to cancel than those who plan to check in on a Sunday. Furthermore, a set of dummy variables designed to capture the booking day is jointly significant. The year of arrival is usually not significantly different from zero. Furthermore, the Wald test of joint significance shows that room category is also relevant in guests' decisions to cancel, with both large rooms and single-bed rooms facing a higher risk.

Cancellation probability significantly varies across booking channels. The probit estimations using the total sample show that the cancellation rate is the highest for offline bookings of leisure guests and lowest for those arranged via travel agencies which is the reference group (see Table 4 in appendix). Based on the total sample of 233,000 bookings, individual leisure guests who booked offline are 12 percentage points more prone to cancelling compared to bookings placed through travel agencies after controlling for individual and booking characteristics. The corresponding marginal effects for OTAs and business guests who booked offline are 10 and 6 percentage points, respectively. These findings do confirm our theoretical prediction that the cancellation rate for online bookings is higher than those for travel agencies. The findings also indicate that the gap in the cancellation probability between online (OTA) and offline booking disappears once the control variables are taken into account, which leads us to reject hypothesis H1.

In addition to including several robustness checks, this paper offers a number of further analytical insights. First, estimations conducted on the snow season (November to April) and each individual arrival month reveal that snow depth on the arrival day is only significant for December and April, and that the magnitude of the relationship is rather small. Snow depth affects only a small share of bookings, which is why it was not included in the final specification. Second, more elaborate models such as the mixed-effects logit and probit models are employed (Rabe-Hesketh and Skrondal, 2012). This makes it possible to account for heterogeneity in the parameters across hotel establishments, seasons, arrival months and/or country of origin (or combinations of the group variables). Results show the magnitude and significance of the marginal effects of the key variables are not sensitive to the more general estimation method. Third, we have checked whether the cancellation of prominent sporting events (such as the FIS Ski World Cup in Levi) due to the lack of snow in 2011 and 2015 played a role in the bookings cancelled at a hotel near Levi, but found no significant relationship. Fourth, we have included bookings that were cancelled on the arrival day, in spite of this being an unrealistic scenario in practice. In any case, these results show that our findings are not sensitive to the inclusion of these cancellations. Finally, we have included measures of consumer sentiment (measured by the consumer confidence indicator published at www.stat.fi), which can stimulate decisions to travel (Dragouni et al., 2016) and thus have an indirect impact on cancellation behaviour. However, consumer sentiment in the main visitor countries had no impact on cancellation behaviour during the period observed.

6 Discussion and conclusions

This study provides new empirical evidence of the probability to cancel a hotel booking, based on unique data originating directly from a booking system. The main results of the probit estimations largely follow the hypotheses derived from the eclectic theoretical approach, where the likelihood of cancellations is significantly related to booking lead time, country of residence, season as well as the composition and size of the of travel group. The timing of booking is crucial, with early bookings exhibiting a significantly higher cancellation probability. Another factor of importance is the channel, where those who book via travel agencies more frequently keep their reservations than those who book online or privately offline. The role of booking lead time and country of residence of the presumptive hotel guest in determining the cancellation probability is more pronounced for online than for offline or travel agency bookings.

These results have several theoretical and practical implications. Theoretical models of cancellation should account for the wide heterogeneity in behaviour among clients. However, a comprehensive theory of the cancellation behaviour still needs to be developed. Managers and practitioners may pay specific attention to those bookings identified as particularly risky, by modifying the cancellation fee upwards, for instance. Consequently, accepting a large number of high risk bookings enables hotels to engage in overbooking. Information on the cancellation probability is also useful for enquiries about available rooms, and possible waiting lists.

This study is subject to several limitations related to the nature of the data at hand. It is based on the booking records of nine hotels that belong to one specific chain. Therefore, the empirical results cannot be generalised to any other group of hotels. In addition, individual characteristics such as gender and information on previous stays in the same hotel might also affect the cancellation behaviour.

Future research could include analysis of information on cancellations with or without a cost, the reasons for cancellations, prior cancellation behaviour of the clients, and cancellation dates. We thus recommend that hotels pay specific attention to features that incorporate useful information for analyses of cancellation behaviour when commercial hotel booking systems are updated or acquired. Gender and previous stays could also be taken into account, although this requires information on the names of guests and their previous bookings.

Literature

Ajzen, I. (1991), "The theory of planned behavior", *Organizational behavior and human decision processes*, 50 No. 2, 179-211.

Antonio, N., de Almeida, A. and Nunes, L. (2017), "Predicting hotel booking cancellations to decrease uncertainty and increase revenue", *Tourism & Management Studies*, Vol. 13, No. 2, pp. 25-39.

Barber, B. M. and Odean, T. (2002), "Online investors: Do the slow die first", *Review of Financial Studies,* Vol. 15 No. 2, pp. 455-488.

Bollen, K. A. and Hoyle, R. H. (1990), "Perceived cohesion: A conceptual and empirical examination", *Social Forces*, Vol. 69 No. 2, pp. 479-504.

Castillo-Manzano, J. I. and López-Valpuesta, L. (2010), "The decline of the traditional travel agent model", *Transportation Research Part E,* Vol. 46, pp. 639-649.

Chen, C. C., Schwartz, Z. and Vargas, P. (2011), "The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers", *International Journal of Hospitality Management,* Vol. 30 No. 1, pp. 129-135.

Chen, C. C. and Xie, K. L. (2013), "Differentiation of cancellation policies in the US hotel industry", *International Journal of Hospitality Management*, Vol. 34, pp. 66-72.

Cheron, E. J. and Ritchie, J. R. B. (1982), "Leisure activities and perceived risk", *Journal of Leisure Research*, Vol. 14 No. 2, pp. 139-54.

Chew, E. Y. T. and Jahari, S. A. (2014), "Destination image as a mediator between perceived risks and revisit intention: A case of post-disaster Japan", *Tourism Management*, Vol. 40, pp. 382-393.

Cheyne, J., Downes, M. and Legg, S. (2006), "Travel agent vs. internet: what influences travel consumer choices?", *Journal of Vacation Marketing,* Vol. 12 No. 1, pp. 41-57.

DeKay, F., Yates, B. and Toh, R. S. (2004), "Non-performance penalties in the hotel industry", *Hospitality Management,* Vol. 23, pp. 273-286.

Delgado, P. (2016), "Cancellations shooting up: implications, costs and how to reduce them", available at: https://www.mirai.com/blog/cancellations-shooting-up-implications-costs-and-how-to-reduce-them/, (accessed March 2017).

Dolnicar, S. and Laesser, C. (2007), "Travel agency marketing strategy: Insights from Switzerland", *Journal of Travel Research,* Vol. 46 No. 2, pp. 133-146.

Donthu, N. and Garcia, A. (1999), "The internet shopper", *Journal of Advertising Research*, Vol. 39 No. 3, pp. 52-52.

Dragouni, M., Filis, G., Gavriilidis, K. and Santamaria, P. (2016), "Sentiment, mood and outbound tourism demand", *Annals of Tourism Research*, Vol. 60, pp. 80-96.

Falk, M and Vieru, M. (in press), "Myth of early booking gains", *Journal of Revenue and Pricing Management*, DOI: 10.1057/s41272-017-0134-9.

Fesenmaier, D. R. and Jeng, J-M. (2000), "Assessing structure in the pleasure trip planning process", *Tourism Analysis*, Vol. 5 No. 1. pp. 13-27.

Garrow, L. A. and Koppelman, F. S. (2004), "Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information", *Journal of Air Transport Management*, Vol. 10 No. 6, pp. 401-411.

Hajibaba, H., Boztuğ, Y. and Dolnicar, S. (2016), "Preventing tourists from cancelling in times of crises", *Annals of Tourism Research,* Vol. 60, pp. 48-62.

Hajibaba, H., Gretzel, U., Leisch, F. and Dolnicar, S. (2015), "Crisis-resistant tourists", *Annals of Tourism Research*, Vol. 53, pp. 46-60.

Hofstede, G. (2001), "*Culture's Consequences. Comparing Values, Behaviors, Institutions, and Organizations Across Nations*", 2nd ed. Thousand Oaks, CA: Sage.

Höpken, W., Fuchs, M., Keil, D. and Lexhagen, M. (2015), "Business intelligence for cross-process knowledge extraction at tourism destinations", *Journal of Information Technology and Tourism*, Vol. 15 No. 2, pp. 101-130.

Höpken, W., Fuchs, M. and Lexhagen, M. (2014), "The knowledge destination – applying methods of business intelligence to tourism destinations", In Wang, J. (ed.) *Encyclopaedia of Business Analytics and Optimization*, IGI Global, pp. 2452-2556.

Iliescu, D.C., Garrow, L.A. and Parker, R.A., (2008), "A hazard model of US airline passengers' refund and exchange behavior", *Transportation Research Part* B Vol. 42 No. 3, pp. 229-242.

Kim, W.G., Cho, M., Kim, D. and Shin, G-C. (2014), "The effect of price dispersion on hotel performance", *Tourism Economics,* Vol. 20 No. 6, pp. 1159-1179.

Koide, T. and Ishii, H. (2005), "The hotel yield management with two types of room prices, overbooking and cancellations", *International Journal of Production Economics,* Vol. 93, pp. 417-428.

Kozak, M., Crotts, J. C. and Law, R. (2007), "The impact of the perception of risk on international travellers", *International Journal of Tourism Research*, Vol. 9 No. 4, pp. 233-242.

LaRose, R. and Eastin, M. S. (2002), "Is online buying out of control? Electronic commerce and consumer self-regulation", *Journal of Broadcasting & Electronic Media,* Vol. 46 No. 4, pp. 549-564.

Law, R. (2006), "The perceived impact of risks on travel decisions", *International Journal of Tourism Research*, Vol. 8 No. 4, pp. 289-300.

Law, R., Leung, K. and Wong, J. (2004), "The impact of the internet on travel agencies", *International Journal of Contemporary Hospitality Management,* Vol. 16 No. 2, pp. 100-107.

Lee, H., Guillet, B. D. and R. Law, R. (2013), "An examination of the relationship between online travel agents and hotels: a case study of choice hotels international and Expedia.com", *Cornell Hospitality Quarterly,* Vol. 54 No. 1, pp. 95-107.

Lepp, A. and Gibson, H. (2003), "Tourist roles, perceived risk and international tourism", *Annals of Tourism Research*, Vol. 30 No. 3, pp. 606-624.

Liu, P. H. (2004), "Hotel demand/cancellation analysis and estimation of unconstrained demand using statistical methods", In I. Yeoman, U. McMahon-Beattie, (Ed). *Revenue Management and Pricing: Case Studies and Applications*. London: Thomson.

Masiero, L., Nicolau, J. L. and Law, R. (2015), "A demand-driven analysis of tourist accommodation price: A quantile regression of room bookings", *International Journal of Hospitality Management*, Vol. 50, pp. 1-8.

Milliken, F. and Martins, L. L. (1996), "Searching for common threads: Understanding the multiple effects of diversity in organizational groups", *Academy of Management Review,* Vol. 21 No. 2, pp. 402-433.

Morales, D. and Wang, J. (2010), Forecasting cancellation rates for services booking revenue management using data mining", *European Journal of Operational Research*, Vol. 202 No. 2, pp. 554-562.

Moutinho, L. (1987), "Consumer behaviour in tourism", *European Journal of Marketing*, Vol. 21 No. 10, pp. 5-44.

Ng, S. I., Lee, J. A. and Soutar, G. N. (2007), "Tourists' intention to visit a country: the impact of cultural distance", *Tourism Management*, Vol. 28 No. 6, pp. 1497-1506.

Nicolau, J. L. and Masiero, L. (2017), "Determinants of advanced booking", *Annals of Tourism Research*, 67, 78-82.

Park, J. Y. and Jang, S. S. (2014), "Sunk costs and travel cancellation: Focusing on temporal cost", *Tourism Management,* Vol. 40, pp. 425-435.

Park, S. and Fesenmaier, D. R. (2014), "Travel decision flexibility", *Tourism Analysis,* Vol. 19, pp. 35-49.

Rabe-Hesketh, S. and Skrondal, A. (2012), *Multilevel and Longitudinal Modeling Using Stata*. 3rd ed. College Station, TX: Stata Press.

Roehl, W.S. and Fesenmaier, D.R. (1992), "Risk perceptions and pleasure travel: An exploratory analysis", *Journal of Travel Research*, Vol. 30 No. 4, pp. 17-26.

Seddighi, H., Nuttall, M. and Theocharous, A. (2001), “Does cultural background of tourists influence the destination choice? An empirical study with special reference to political instability”, *Tourism Management,* Vol. 22 No. 2, pp. 181-191.

Sharifpour, M., Walters, G., Ritchie, B. and Winter, C. (2014), “Investigating the role of prior knowledge in tourist decision making: A structural equation model of risk perceptions and information search”, *Journal of Travel Research*, Vol. 53 No. 3, pp. 307–322.

Sirakaya, E. and Woodside, A. G. (2005), “Building and testing theories of decision making by travellers”, *Tourism Management,* Vol. 26, pp. 815-832.

So, S. I. and Lehto, X. Y. (2007), “The situational influence of travel group composition”, *Journal of Travel & Tourism Marketing*, Vol. 20 No. 3-4, pp. 79-91.

Subramanian, J., Jr, Stidham, S. and Lautenbacher, C. J., (1999), “Airline yield management with overbooking, cancellations, and no-shows”, *Transportation Science,* Vol. 33 No. 2, pp. 147-167.

Thornton, P. R., Shaw, G. and Williams, A. M. (1997), “Tourist group holiday decision making and behaviour: The influence of children”, *Tourism Management*, Vol. 18 No. 5, pp. 287-297.

Toh, R. S. (1986), “Coping with no-show, late cancellations and oversales: American hotels out-do the airlines”, *International Journal Hospitality Management,* Vol. 5 No. 3, pp. 121-125.

Tse, S. M. T. and Poon, Y. T. (2017), “Modeling no-shows, cancellations, overbooking, and walk-ins in restaurant revenue management”, *Journal of Foodservice Business Research*, Vol. 20 No. 2, pp. 127-145.

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT press.

Xiong, J. and Hansen, M. (2013), “Modelling airline flight cancellation decisions”, *Transportation Research Part E: Logistics and Transportation Review,* Vol. 56, pp. 64-80.

Zakhary, A., Atiya, A. F., El-Shishiny, H. and Gayar, N. E. (2011), “Forecasting hotel arrivals and occupancy using Monte Carlo simulation”, *Journal of Revenue and Pricing Management*, Vol. 10 No. 4, pp. 344-366.

Appendices

Table 1. Descriptive statistics of distribution of bookings and cancellation rate by type of booking channel

	Bookings incl. cancellations (%)			Cancellation rate (%)		
	Offline[a]	Travel agency[a]	OTAs[a]	Offline	Travel agency	OTAs
			Booking lead time			
Booking 1-4 days before arrival	20.3	7.5	19.6	8.9	2.9	2.9
Booking 5-9 days before arrival	14.7	12.2	11.9	9.2	2.3	4.8
Booking 10-24 days before arrival	28.7	38.5	22.5	10.0	2.5	11.0
Booking 25-49 days before arrival	17.5	28.9	19.4	13.9	4.2	20.8
Booking 50-99 days before arrival	11.2	7.9	15.5	16.8	8.0	29.2
Booking 100+ before arrival	7.6	4.9	11.1	19.5	12.8	39.5
			Season			
Non-holiday season	79.5	62.6	65.8	11.2	3.7	15.5
Christmas holidays	2.2	5.6	8.5	16.9	6.4	27.8
Winter holidays	9.8	8.4	12.3	15.9	4.8	18.5
Easter holidays	3.0	0.5	4.4	17.7	6.3	18.0
Summer holidays	5.5	22.8	9.1	8.3	3.7	9.7
			Number of adult guests			
No. of adult guests= 1	49.5	49.5	22.8	12.3	3.8	11.9
No. of adult guests= 2	48.1	48.1	56.5	10.2	5.4	14.6
No. of adult guests= 3	1.7	1.7	8.6	10.2	3.7	21.4
No. of adult guests= 4+	0.7	0.7	12.1	12.5	3.6	24.8
			Number of planned nights			
No. of nights= 1	34.6	56.2	39.5	10.3	3.6	12.5
No. of nights= 2	23.6	9.5	21.4	13.8	4.1	15.5
No. of nights= 3	14.6	11.4	15.4	17.0	7.5	20.1
No. of nights= 4	8.0	6.8	8.2	18.7	9.5	26.5
No. of nights= 5+	19.2	16.1	15.5	20.8	5.5	27.1
			Children			
Yes	5.0	8.9	2.2	13.7	2.3	9.0
No	95.0	91.1	97.8	11.5	4.1	16.5
			Country of residence			
Belgium	0.3	0.5	0.6	9.4	4.1	17.9
China	0.4	0.5	1.3	8.3	6.3	37.6
Denmark	0.1	0.8	0.2	20.9	2.7	16.1
Finland	79.4	10.3	54.9	11.1	5.4	12.3
France	0.6	10.8	1.9	7.1	2.0	17.5
Germany	5.3	33.8	3.0	25.9	3.6	21.0
Great Britain	2.7	11.3	2.5	5.7	1.7	21.4
Hong Kong, Singapore, Taiwan	0.0	2.9	1.3	11.1	2.1	29.9
Italy	0.5	2.7	1.3	11.1	6.0	24.7
Japan	0.3	7.1	4.7	19.4	7.1	16.7
Netherlands	1.0	4.4	0.9	4.2	4.6	16.3
Norway	2.2	0.6	6.0	13.6	8.0	11.6
Other Non-Europe	0.7	1.4	1.8	10.3	7.0	25.0
Other Europe	1.9	4.4	2.5	16.1	3.9	21.9
Poland	0.3	0.7	0.2	7.4	5.2	26.6
Russia	1.5	1.1	9.3	14.6	6.7	31.4
Spain	0.4	1.2	1.4	21.6	7.4	28.5
Sweden	1.4	0.5	4.2	8.6	4.8	17.4
Switzerland	0.6	4.8	1.5	9.5	4.2	9.6
United States	0.4	0.1	0.5	3.9	6.3	17.8

Note: [a] For each set of variables the share of booking categories sum up to 100 per cent.

Table 2. Probit model estimates on the factors affecting the probability of cancellations by type of channel

	Offline		Travel agency		OTAs	
	dF/dx	z -value	dF/dx	z -value	dF/dx	z-value
Booking 5-9 days before arrival (ref.1-4)	0.017***	4.43	-0.006***	-2.66	0.053***	4.51
Booking 10-24 days before arrival	0.032***	9.50	-0.005**	-2.37	0.164***	15.70
Booking 25-49 days before arrival	0.056***	14.30	0.017***	7.04	0.283***	24.64
Booking 50-99 days before arrival	0.088***	18.89	0.063***	16.13	0.381***	29.88
Booking 100+ before arrival	0.107***	19.40	0.109***	21.72	0.519***	35.66
Christmas holidays (ref. non-holiday)	0.023***	2.82	0.009***	3.59	-0.006	-0.72
Winter holidays	0.012***	2.98	0.000	0.11	0.005	0.63
Easter holidays	0.025***	4.00	0.008	1.06	0.021*	1.86
Summer holidays	-0.016*	-1.82	0.006***	2.94	-0.009	-0.59
No. of adult guests= 2 (ref=1)	0.003	1.30	0.000	-0.07	-0.022***	-3.79
No. of adult guests = 3	0.022***	3.71	0.016***	4.07	0.014	1.47
No. of adult guests = 4+	0.045***	6.85	0.028***	4.26	0.016*	1.66
No. of nights= 2 (ref=1)	-0.024***	-8.34	-0.005***	-2.83	0.005	0.90
No. of nights= 3	-0.037***	-11.71	-0.012***	-6.86	0.029***	4.35
No. of nights= 4	-0.029***	-7.47	-0.011***	-4.78	0.042***	5.01
No. of nights= 5+	-0.030***	-9.79	-0.010***	-5.14	0.041***	5.57
Guests with children	-0.002	-0.42	-0.008***	-3.37	-0.068***	-5.41
Belgium (ref. Finland)	0.011	0.56	-0.022***	-5.72	-0.031	-1.34
China	-0.006	-0.33	-0.011**	-2.10	0.240***	11.00
Denmark	0.123***	3.70	-0.018***	-4.40	0.013	0.31
France	-0.036***	-2.77	-0.027***	-15.69	-0.011	-0.78
Germany	0.111***	19.84	-0.017***	-9.29	0.002	0.15
United Kingdom	-0.065***	-10.42	-0.030***	-17.05	-0.004	-0.29
Hong Kong, Singapore, Taiwan	-0.033	-0.74	-0.027***	-12.60	0.061***	3.33
Italy	0.004	0.29	-0.008***	-3.14	0.053***	2.91
Japan	0.098***	4.62	-0.009***	-4.42	-0.005	-0.50
The Netherlands	-0.053***	-4.81	-0.016***	-7.97	-0.035*	-1.85
Norway	0.041***	5.29	0.006	1.00	-0.008	-0.84
Other Non-Europe	0.012	0.88	-0.008**	-2.35	0.033**	2.23
Other Europe	0.064***	7.45	-0.017***	-7.81	0.031**	2.31
Poland	0.006	0.26	-0.008*	-1.70	0.054	1.24
Russia	0.054***	5.55	-0.004	-1.01	0.148***	15.56
Spain	0.104***	6.02	-0.008**	-2.15	0.088***	4.63
Sweden	-0.022**	-2.48	-0.010*	-1.74	0.034***	3.10
Switzerland	-0.021	-1.58	-0.021***	-11.27	-0.071***	-5.01
United States	-0.055***	-2.99	-0.015*	-1.81	0.029	1.05
Number of observations	87,727		118,203		27,191	
Pseudo R^2	0.07		0.09		0.17	
Correctly classified, %	92		92		91	

Notes: Asterisks ***, **, * denote significance at the 1%, 5% and 10% level, respectively. Table reports the marginal effects, dF/dx, and the corresponding z values. Probit model is estimated by maximum likelihood with standard errors clustered across hotels season pairs (winter season and no winter season).

Table 3. Probit model estimates of the chosen control variables by type of channel

	Offline		Travel agency		OTAs	
	dF/dx	z-value	dF/dx	z-value	dF/dx	z-value
Business guests (ref. non-business)	-0.052 ***	-20.08				
Associations	-0.067 ***	-20.47				
Arrival month =Feb (ref. Jan)	-0.008	-1.61	-0.003	-1.41	0.010	0.93
Arrival month =Mar	-0.006	-1.45	-0.004 *	-1.79	-0.002	-0.27
Arrival month =Apr	-0.015 ***	-3.09	-0.010 ***	-2.61	-0.035 ***	-3.84
Arrival month =May	0.021 *	1.82	-0.017 ***	-4.83	0.020	0.57
Arrival month =Jun	-0.018 ***	-2.57	-0.020 ***	-8.94	-0.027 *	-1.92
Arrival month =Jul	-0.047 ***	-5.29	-0.022 ***	-8.56	-0.016	-0.91
Arrival month =Aug	-0.042 ***	-8.23	-0.019 ***	-8.73	-0.027 **	-2.56
Arrival month =Sep	-0.040 ***	-9.03	-0.023 ***	-11.46	-0.037 ***	-3.93
Arrival month =Oct	-0.008	-1.05	-0.010 **	-2.12	0.032 *	1.79
Arrival month =Nov	-0.019 ***	-3.52	0.007 *	1.83	0.014	1.20
Arrival month =Dec	0.009	1.48	-0.001	-0.34	0.032 ***	3.20
Arrival day Monday (ref. Sunday)	-0.002	-0.22	0.001	0.57	0.007	0.74
Tuesday	-0.010	-0.94	0.000	-0.08	0.016	1.14
Wednesday	0.003	0.34	-0.002	-0.70	0.008	0.77
Thursday	0.004	0.43	0.002	0.58	0.028 ***	2.71
Friday	0.007	0.74	0.000	0.06	0.019 *	1.92
Saturday	-0.003	-0.34	-0.005 *	-1.65	0.005	0.63
Wald-tests (p-values):						
Dummy variables Arrival day	0.10		0.00		0.00	
Dummy variables Booking day	0.04		0.00		0.00	
Dummy variables Arrival year	0.81		0.00		0.05	
Dummy variables Hotel	0.00		0.00		0.00	
Dummy variables Room category	0.00		0.00		0.00	

Notes: Asterisks ***, **, * denote significance at the 1%, 5% and 10% level, respectively. Table reports the marginal effects, dF/dx, and the z-value. A set of dummy variables measuring the arrival day, booking day, arrival year, hotel and room category are included but not reported due to space limitations.

Table 4. Probability of cancellations (all booking channels).

	dF/dx		z-value
Offline Individual (ref Travel agency)	0.126	***	6.65
Offline Business	0.061	***	3.76
OTAs	0.103	***	5.40
Booking 5-9 days before arrival (ref. 1-4)	-0.008		-0.70
Booking 10-24 days before arrival	0.058	**	2.31
Booking 25-49 days before arrival	0.017		0.98
Booking 50-99 days before arrival	-0.030	***	-4.14
Booking 100+ before arrival	0.024		1.20
Christmas holidays (ref. non-holiday)	-0.038	***	-6.71
Winter holidays	-0.017		-1.16
Easter holidays	0.030		1.95
Summer holidays	0.023	**	2.38
No. of adult guests = 2 (ref=1)	-0.005		-0.49
No. of adult guests = 3	0.011	**	2.29
No. of adult guests = 4+	0.022	**	2.17
No. of nights= 2 (ref=1)	0.019		1.04
No. of nights= 3	0.036	**	2.49
No. of nights= 4	0.057	***	4.73
No. of nights= 5+	0.046	**	2.10
Guests with children	-0.002		-0.18
Belgium (ref. Finland)	-0.011		-1.04
China	-0.014	**	-2.16
Denmark	0.006	***	2.54
France	0.000		-0.04
Germany	0.002		0.56
United Kingdom	0.006	*	1.71
Hong Kong, Singapore, Taiwan	0.004		0.83
Italy	-0.006		-1.61
Japan	0.000		-0.09
The Netherlands	-0.001		-0.26
Norway	0.000		-0.10
Other Non-Europe	-0.002		-0.70
Other Europe	0.003		0.86
Poland	0.002		0.34
Russia	0.010	***	2.67
Spain	0.008	**	2.22
Sweden	0.019	***	3.42
Switzerland	0.008		1.28
United States	0.009	*	1.69
Control variables	yes		
Wald-tests (p-values):			
Dummy variables Arrival day	0.00		
Dummy variables Booking day	0.30		
Dummy variables Arrival year	0.00		
Dummy variables Hotel	0.00		
Dummy variables Room category	0.00		
Number of observations	233121		
Pseudo R^2	0.12		
Correctly classified, %	92		

Notes: Asterisks ***, **, * denote significance at the 1%, 5% and 10% level, respectively. Table reports the marginal effects, dF/dx, and the z-value.