

Giannoni, Sauveur; Brunstein, Daniel; Guéniot, Florian; Jouve, Johan

Article

Multichannel distribution strategy of Airbnb hosts

Annals of Tourism Research Empirical Insights

Provided in Cooperation with:

Elsevier

Suggested Citation: Giannoni, Sauveur; Brunstein, Daniel; Guéniot, Florian; Jouve, Johan (2021) : Multichannel distribution strategy of Airbnb hosts, Annals of Tourism Research Empirical Insights, ISSN 2666-9579, Elsevier, Amsterdam, Vol. 2, Iss. 1, pp. 1-10, <https://doi.org/10.1016/j.annale.2021.100017>

This Version is available at:

<https://hdl.handle.net/10419/337680>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

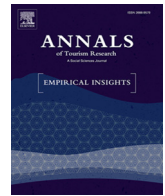


<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Contents lists available at ScienceDirect

Annals of Tourism Research Empirical Insights

journal homepage: www.elsevier.com/locate/annale

Multichannel distribution strategy of Airbnb hosts

Sauveur Giannoni ^{*}, Daniel Brunstein, Florian Guéniot, Johan Jouve

UMR CNRS 6240 LISA – Université de Corse, Avenue Jean Nicoli, 20250 Corte, Corse, France

ARTICLE INFO

Article history:

Received 17 September 2020
 Received in revised form 12 April 2021
 Accepted 22 April 2021

Keywords:

Multichannel
 Home-sharing
 Airbnb
 Zero-inflated models
 Endogeneity

ABSTRACT

The literature on the peer-to-peer hospitality market largely assumes that Airbnb hosts are loyal to the platform. However, similar to hoteliers, Airbnb hosts have access to various platforms to diversify their distribution channels. In cases where hosts significantly diversify their distribution channels, they should be considered rather as professional competitors mimicking hoteliers. Proposing an original image-recognition approach to obtain a proxy of the number of platforms used by Airbnb hosts, this paper assesses the probability that a home-sharer practices multichannel distribution. Drawing on a sample of more than 3900 hosts from the region of Corsica in France, we show that a multichannel distribution strategy is commonly adopted. Furthermore, a significant fraction of small hosts uses three or more channels.

1. Introduction

Online vacation rental platforms emerged in the mid-1990s and multiplied throughout the 2000s. In 2008, Airbnb was founded based on the concept of a user community and grew exponentially worldwide to become the leading player in the short-term rental market. Booking.com and Expedia have spearheaded worldwide online travel agency development through the buyout of smaller competing short-term rental platforms. Market concentration over the last 15 years has reduced the number of players, with five global heavyweights (e.g. Airbnb, Booking.com, VRBO, TripAdvisor, and Tujia) accounting for more than three quarters of global market revenues, according to the blog Skift (Geerts, 2019). Nevertheless, many smaller regional and local actors have resisted. Hence, hosts practising short-term rental have various competing online platforms to use: some are global players such as Airbnb, VRBO, and Booking.com; others are only present in certain regional markets such as Interchalet in Europe; and, finally, national or local players operate in many markets.

In the traditional hospitality sector, as pointed out by several authors interested in marketing strategies in the traditional hospitality sector (Beritelli & Schegg, 2016; Stangl, Inversini, & Schegg, 2016), multichannel distribution is a widespread strategy. Hotels simultaneously use several distribution channels, physical or virtual, to increase their visibility and improve their competitiveness. Some sellers (hosts) on home-sharing platforms are likely to mimic such behaviour to enhance their performances. Notably, factual evidence of the widespread use of multiple competing platforms by rental owners or managers is, in part, directly observable. In the market for services to owners and managers, a multitude

of products and auxiliary services exist to facilitate the distribution of listings on multiple competing platforms. For example, the professional blog Padlifter lists approximately 50 applications that allow multichannel management (Padlifter, 2021). These tools offer solutions to the pitfalls that the host experiences when renting an asset on multiple platforms. Most of these applications provide a channel manager to organise ‘multiple channels in one place’. This feature allows users to synchronise calendar availability, reservations, photos, and information, not only with major distribution sites such as Airbnb, Booking.com, and VRBO but also on regional or local platforms.

The use of multichannel marketing in the short-term rental sector could affect various issues addressed by the research. The multichannel strategy can be a differentiation tool in markets in which there is significant competition between hosts. This strategy could also improve the knowledge of hosts’ behaviours, particularly regarding their loyalty to the platforms they use, and their know-how, experience, and level of professionalism. Despite the practical relevance of this issue, academic research on Airbnb has never attempted to quantify the simultaneous use of platforms by hosts. Furthermore, several authors (Ke, 2017; Krause & Aschwanden, 2020; Li, Moreno, & Zhang, 2016) have demonstrated that the home-sharing market comprises non-professional players and professional players. Because the objective of these types of hosts differs, their online distribution strategy is likely to vary considerably. In summary, this paper addresses two related questions. Is the use of several online platforms, namely, distribution channels, frequent behaviour for hosts in the online short-term rental market? Do strategies and intensity of use vary according to their level of professionalism?

^{*} Corresponding author.

E-mail addresses: giannoni_s@univ-corse.fr (S. Giannoni), brunstein_d@univ-corse.fr (D. Brunstein), gueniot_f@univ-corse.fr (F. Guéniot), jouve_j@univ-corse.fr (J. Jouve).

This paper aims to assess the probability of a host practising multichannel distribution. We exploit a sample of 3916 Airbnb hosts operating in Corsica. We posit that Corsica is an appropriate case study because, for example, in 2014, several rental platforms and many local players (i.e. travel agencies, real estate agencies) operated the market, and some remain present today. Back then, the leading operator in Corsica was VRBO; Airbnb marketed 2000 listings of the 26,594 known listings at the time. In 2017, for entire homes only, Airbnb marketed 9734 listings. The same year, 36% of Airbnb hosts also used VRBO. This historical perspective could support the idea that VRBO hosts took advantage of the emergence of Airbnb to increase their audience without leaving their original platform. To assess the extent of multichannel practice, we use an original image-recognition approach to obtain a proxy of the number of distribution channels used by Airbnb hosts. The use of image recognition raises a number of methodological issues related to the data collection process and potential omitted variables bias. An adequate methodological design is tailored by combining inflated-zero models and a two-stage residual inclusion method to manage data collection issues and potential endogeneity, respectively.

We show that the probability for a host to use Airbnb and at least one additional online distribution channel is above 51%. Moreover, we show that the number of competing platforms used depends on the level of professionalism of the hosts and that multichannel marketing is a strategy adopted by a significant proportion of hosts marketing one or two properties. Drawing on a sample of more than 3900 hosts from the region of Corsica in France, we estimate appropriate count models and derive the probability for the types of hosts to use a multichannel approach and, in this manner, mimic hospitality firms. We also establish that hosts with many properties for rent are the most likely to use many distribution channels.

The next section reviews the literature on multichannel distribution in hospitality and discusses whether hosts are hospitality professionals. Sections 3 and 4 detail the data and econometric strategy. Section 5 presents our results, and Section 6 discusses our findings, draws conclusions, and raises questions for future research.

2. Literature review

The first section reviews the academic literature on the Airbnb hosts' motivations and their level of professionalism. The second section presents the studies and findings on the use of a multichannel strategy in the traditional hospitality sector.

2.1. Hosts' Motivation and Professionalism

Hosts' motivations are often related to both the financial and the social benefits of the experience (Ikkala & Lampinen, 2015). Financial considerations remain the main motivation for homeowners renting their homes on a platform (Karlsson & Dolnicar, 2016; Visser, Erasmus, & Miller, 2017). However, financial motivations can differ. Motivations may be an income supplement for coping with a particular situation or a regular income to support a loan or rent payment (Lampinen & Cheshire, 2016). For example, hosts motivated by economic gain are often not on Airbnb to make large profits but use the platform to cover their fixed costs. Others desperately need money to pay their bills (Gibbs, Guttentag, Gretzel, Morton, and Goodwill, 2018; Gibbs, Guttentag, Gretzel, Yao, and Morton, 2018; Karlsson & Dolnicar, 2016). Airbnb hosts have the possibility to play on several parameters (price, obtaining the Superhost badge, use of instant booking reservation mode, specific cancellation policy) to maximise their income (Benítez-Aurioles, 2018; Gibbs, Guttentag, Gretzel, Morton, and Goodwill, 2018; Gibbs, Guttentag, Gretzel, Yao, and Morton, 2018; Kwok & Xie, 2019).

Adopting multichannel distribution could be an additional strategy for hosts to maximise their profits by broadening their customer base. A legitimate thought would be that a multichannel strategy is a step toward professional behaviour, requiring additional skills and experience. Authors' interest in hosts' level of professionalism has shown that similar to other

markets of the 'sharing economy', the supply side often comprises professional (i.e. experienced) players practising a 'profit-oriented supply' approach and non-professional (i.e. inexperienced) players with a more occasional activity (Ke, 2017; Krause & Aschwanden, 2020; Li et al., 2016).

A commonly used indicator of professional behaviour on an Airbnb market is the number of listings per host. In the majority of available market studies, a host is considered a professional when he/she has more than one listing (CBRE, 2017; Dredge, Gyimóthy, Birkbak, Jensen, & Madsen, 2016; Gurran & Phibbs, 2017; Li et al., 2016). Other studies apply a threshold of two or three listings (Schneiderman, 2014). In a global Airbnb market study, Ke (2017) shows that there is no optimal threshold that enables differentiation between professional and non-professional hosts. He argues that the use of a specific threshold would bias the perception of distribution as a whole and could be misleading. Ke then studied how various measures of interest change along with the number of listings per host. This strategy is often used by authors working on host specifications and behaviour (Gibbs, Guttentag, Gretzel, Morton, and Goodwill, 2018; Gibbs, Guttentag, Gretzel, Yao, and Morton, 2018; Kwok & Xie, 2019; Li et al., 2016; Magno, Cassia, & Ugolini, 2018; Oskam, van der Rest, & Telkamp, 2018; Wang & Nicolau, 2017).

2.2. Multichannel distribution in traditional hospitality

The tourism and hospitality sector has substantially benefited from various technological advances. The advent of the internet in the 1990s resulted in an unprecedented revolution in the hospitality and tourism industries (Law, Buhalis, & Cobanoglu, 2014). One result of this technological disruption was the creation of online travel agencies, which offer travel products such as airline tickets (e.g. Kayak, Opodo) and hotel rooms (e.g. Booking.com, Expedia.com, TripAdvisor) from multiple suppliers directly to consumers. The arrival of online travel agencies has further amplified the diversification of distribution channels for professionals (Gazzoli, Kim, & Palakurthi, 2008). The structure of the tourism industry has taken the form of a complex global network (Kracht & Wang, 2010).

From the suppliers' perspective, this revolution has enabled them not only to increase their sales but also to access new marketing opportunities, providing them with the opportunity to add value to their products. Multichannel distribution in the hospitality sector now includes four types of channels simultaneously used by professionals in the sector (Stangl et al., 2016): (i) physical means of reservation (e.g. telephone, fax, letter, and walk-ins), (ii) real-time channels (e.g. travel platforms), (iii) electronic channels (e.g. email, electronic forms), and (iv) online travel agencies. Today, through their respective brands, hoteliers have, for the most part, joined the main groups sharing the global online travel agency market (Beritelli & Schegg, 2016). For example, according to Beritelli and Schegg (2016), hoteliers in Switzerland, Germany, and Austria use an average of 3.6 online travel agencies in addition to physical distribution channels, and the latter still play a predominant role in the distribution of hotel rooms.

In summary, a significant fraction of the hosts attempt to profit, and multichannel distribution is proven to be a common practice to enhance profitability in the accommodation industry. Similar to hoteliers, rational Airbnb hosts may be willing to increase the visibility of their listings to enhance their performance. Despite their empirical relevance, some questions have been ignored because of data unavailability. To which extent do Airbnb hosts use alternative platforms (e.g. Booking, VRBO, TripAdvisor)? Do hosts tend to mimic hospitality firms in using multiple distribution channels? Furthermore, do professional hosts tend to diversify their online channels more than non-professional hosts do? This paper proposes to fill these gaps.

3. Data

This section describes the data used in our study. We first explain how the data on the host's online multichannel behaviour is collected and then describe the sample.

3.1. Multichannel data acquisition

We define this online multichannel practice as a host posting listings on several online rental platforms. Because the main platforms are intermediaries that prohibit a direct relationship between owner and consumer, this study focuses only on distribution channels mediated by an online platform and excludes other distribution channels, such as owner websites, email, or phone.

This paper uses Airbnb data obtained from AirDNA (<http://airdna.co/>), a data analytics company providing scraped data from the Airbnb website. The dataset from AirDNA contains Airbnb listings with the location, amenities, detailed host information, and, most importantly for us, picture used for the main property listing. Data are scraped daily. Ioannides, Röslmaier, and van der Zee (2019) cross-reference AirDNA data for the city of Utrecht and conclude that the listing and host data had been correctly scraped.

To identify Airbnb ads distributed on other platforms, we use image-recognition technology. We hypothesised that the main image of the Airbnb ad would be part of the series of photographs displayed on a competing platform. Image recognition makes it possible to identify the same listing on several platforms. We used Google's Search by Image tool, which analyses the submitted picture and compares it with billions of other images in Google's databases before returning matching results. Google's image database is built with the help of robots that constantly crawl and index the World Wide Web (Google, 2021). Crawling may cause a slowdown of the explored website. Google advises large sites to implement search engine optimisation by helping Google crawlers to index or not index their content. The search engine optimisation strategy implemented by website managers plays on the frequency of indexing and updating. The possibility that the images of a listing marketed on a competing website are not indexed by Google and are not identified is one of the justifications for the choice of our econometric model discussed next.

To be sure that a host is practising multichannel marketing, we need to be sure of simultaneous marketing. To avoid counting listings that would have left one platform for another without multichannelling, we considered that the delay between the date of the dataset (October 2017) and the execution of the image search (Jan 2018) was sufficient for the exploration and indexing to be updated. Notably, when deleting images from a website, updating the indexing takes several weeks to a few months and depends on how often the Google image Googlebot explores and indexes the websites. For the Airbnb ads, we only performed the image search on ads created before October 2017 and still live on the date of the image search (Jan 2018).

For each listing in our AirDNA sample, Google Cloud Vision API, powered by Artificial Intelligence, was used to search only for identical images. This software works even if the picture is re-sized. We must specify that only image recognition is used in our methodology; we have not introduced the textual description of the listings.

The API provides a row set of URLs that are then cleaned up based on the domain names to retain the origin of the data. The URLs not related to rental platforms are removed to keep rental platforms only. The search engine for vacation rental websites to aggregate ads from rental platforms such as Trivago or HomeToGo is excluded too. Finally, the listings associated with the same hosts are aggregated. This cleaning process provides us with the total number of platforms on which each listing is present. In this study, the method was applied to the French island of Corsica (population 320,000), located in the western Mediterranean Sea, where tourism-related consumption accounts for one third of Gross Domestic Product. The existence of a developed short-term rental market has become an economic and political issue, as observed in many places worldwide. Stakeholders in the traditional hospitality sector report the fierce competition with hosts considered shadow-hospitality providers. Back in 2015, 20 short-term rental platforms advertising more than 25,000 listings were identified in Corsica (Marc Simeoni Consulting, 2015) in a report commissioned by the local authority responsible for tourism.

As shown in Table 1, even in a small (but attractive) market, no fewer than 22 platforms are in competition with Airbnb. Of the hosts included in our sample, 35.78% are also present on VRBO, and two individuals also use the German platform Interchalet.

3.2. Description of the sample

The original dataset provided by AirDNA contains outdated hosts and listings because of the lag between the AirDNA scraping date in October 2017 and our reverse image search using Google API conducted in February 2018. Furthermore, we focus on hosts renting entire homes for three reasons. First, this group represents more than 80% of the total listings. Second, we posit that hosts renting rooms are likely to substantially differ in motivation from hosts sharing flats or homes. And last, renting a shared or a private room in an apartment is an experience that is, by nature, different from spending a few days in a traditional hotel (in terms of, e.g. privacy, services, freedom of action). Our final sample is 3916 hosts in the Corsican short-term rental market in October 2017. For each host, variables are created, aggregating listing information. Table 2 describes these variables. Table 3 provides a statistical description of the sample.

The variable *PLATS* is our explanatory variable in the following estimates. *PLATS* is the result of the image-recognition process. This numeric variable takes values from 0 to 5. It corresponds to the number of renting platforms other than Airbnb on which a host is identified. For each listing associated with a given host profile, we know the number of rental platforms on which the same picture is identified. *PLATS* is the highest number of platforms, Airbnb excepted, on which at least one listing associated with the host profile has been identified. For example, if *PLATS* takes the value 2, the host has at least one listing on Airbnb and two other platforms simultaneously. One methodological issue in our study is related to the following: when *PLATS* takes the value 0, the meaning of this value is misleading. The natural interpretation is that the host is on Airbnb only. However, the construction of the variable *PLATS* relies on two important assumptions: the main image used on Airbnb belongs to the set of images used by the host to promote his/her listing on competing platforms, and the Google Vision API results are fully reliable.

The first assumption could be false when the main image used on Airbnb is not used elsewhere, and this decision fully depends on the host. The second assumption could also be false if the Google crawling bots have not indexed the images, as explained. Thus, the count variable *PLATS* is not perfect. Specifically, some 0s observed in the sample occur because one (or both) of the prior assumptions are not verified. Some hosts

Table 1
Platforms identified during the image-search process and percentage of Airbnb hosts using each platform.

Platform name	Website	Hosts %
VRBO	www.vbro.com	35.78%
Booking.com	www.booking.com	11.31%
TripAdvisor	www.tripadvisor.fr	9.55%
Cybevasion	www.cybevasion.fr	9.12%
SeLogger.com	vacances.seloger.com	4.26%
Expedia	www.expedia.com	2.15%
Vivaweek	www.vivaweek.com	2.04%
Vivastreet	www.vivastreet.com	1.84%
PaP	www.papvacances.fr	1.76%
Vacances-Location	www.vacances-location.net	1.56%
Media Lite Group	www.ihacom.co.uk	1.43%
Gites de France	www.gites-de-france.com	1.20%
Mediavacances	www.mediavacances.com	0.84%
Locasun Group	www.locasun.co.uk	0.82%
Windu	www.windu.com	0.72%
Lastminute.com	www.lastminute.com	0.38%
e-domizilGmbH	www.e-domizil.com	0.23%
Vacancesweb	www.vacancesweb.be	0.20%
SantiExpertimmo	www.santi-expertimmo.com	0.15%
TravelFactory	www.locatour.com	0.15%
Roomlala	en-fr.roomlala.com	0.15%
Interchalet	www.interchalet.de	0.05%

Table 2
Description of variables.

Variable name	Variable description
Variables of interest	
PLATS	Number of alternative platforms on which a host is identified
NLISTS	Total number of listings managed
Host type variable (HT_j)	
$HT_1 = 1$	Host manages 1 listing [base level]
$HT_2 = 2$	Host manages 2 listings
$HT_3 = 3$	Host manages more than 2 listings
Cancellation policy variable (CP_n)	
$CP_{FLEX} = 1$	Typical cancellation policy of the host is flexible
$CP_{MODE} = 2$	Typical cancellation policy of the host is moderate
$CP_{STRICT} = 3$	Typical cancellation policy of the host is strict
$CP_{STRICT30} = 4$	Typical cancellation policy of the host is strict 30
$CP_{OCP} = 5$	Other cancellation policy behaviours [base level]
Property type variable (PT_m)	
$PT_{APART} = 1$	Host typically rents apartments
$PT_{HOUSE} = 2$	Host typically rents houses
$PT_{OPT} = 3$	Other rental behaviours [base level]
Control variables	
SUPER (0/1)	Host is labelled SUPER HOST by Airbnb
POOL (0/1)	Host typically provides a pool
ROOMS	Average number of bedrooms per listing on Airbnb
PRICE	Average price per night in a host property
Variable for the inflate model	
PHOTOS	Average number of photos per listing on Airbnb

Table 3
Descriptive statistics.

Variable	Mean	Std. Dev.	Min.	Max.	N
PLATS	0.813	1.005	0	5	3916
HT_1	0.671	0.470	0	1	3916
HT_2	0.180	0.384	0	1	3916
HT_3	0.149	0.356	0	1	3916
CP_{FLEX}	0.184	0.388	0	1	3916
CP_{MODE}	0.086	0.280	0	1	3916
CP_{STRICT}	0.661	0.473	0	1	3916
$CP_{STRICT30}$	0.003	0.058	0	1	3916
CP_{OCP}	0.065	0.247	0	1	3916
PT_{APART}	0.35	0.477	0	1	3916
PT_{HOUSE}	0.32	0.467	0	1	3916
PT_{OPT}	0.33	0.47	0	1	3916
SUPER	0.054	0.226	0	1	3916
POOL	0.259	0.438	0	1	3916
ROOMS	2.010	1.302	0	9	3916
PRICE	214.068	242.387	12	3989	3916
PHOTOS	17.195	10.325	1	128	3916

are always 0 because of the characteristics of the data-generation process. In the next section, we explain how we used zero-inflated count models to consider this technical issue.

HT_j is our explanatory variable of interest and denotes the host type. It is constructed from the variable *NLISTS*, indicating the total number of listings managed by a given host on Airbnb. The aim of this paper is to understand whether hosts are mimicking the behaviour of hospitality firms by offering their listings on several platforms to increase their visibility. We are attempting to verify whether the behaviour of the hosts varies according to their type. We define the variable HT_j as a qualitative variable taking three possible values. HT_j is equal to 1 if the host manages one listing only on Airbnb; it takes the value 2 if the host manages two listings; it takes the value 3 if the host manages more than two listings. Although imperfect, this classification seems natural and is common in the literature. Hosts with one listing on Airbnb are *very small players*. Hosts with two listings are *mid-size players*. Hosts managing three or more listings are *big players*.

We could have estimated a model including the HT_j variable only. However, to control for some specific characteristics of the host, we included additional controls not central to the analysis. For example, the variable *POOL* is a dummy variable used to consider a host's propensity to rent

'luxury' properties. The variable *SUPER* is a dummy variable, taking the value 1 if the host is recognised as a 'Superhost' according to Airbnb standards. *PRICE* and *ROOMS* are respectively the mean number of bedrooms and listed price for one night over all the listings associated with the host. Additional controls help us account for some observed host characteristics (typical cancellation policy, CP_n , or typical property type, CP_n). A behaviour is said to be typical when the host adopts this behaviour or provides this amenity for more than 50% of the managed listings; the underlying idea is that some hosts tend to develop behavioural patterns.

PHOTOS is the average number of photos associated with a listing posted by a host. Although this variable is not used as a control in the main model, it helps us overcome the inflated-zero issue discussed in the section presenting the econometric strategy.

4. Econometric strategy

We use econometric regressions to measure the effect of the host type on the probability of diversifying the portfolio of distribution channels. The image-recognition process provides information on the number of platforms used by each host. As presented in step 1 of our method (see Fig. 1), we use count data models. Our explanatory variable (*PLAT*) is a count variable and even if it is common, using a traditional Ordinary Least Squares approach is not appropriate in this case (Cameron and Trivedi, 2013). Notably, we must manage some methodological issues. We explain how we follow a three-step procedure to select our final specification; we consider both specific constraints raised by the use of image recognition and in managing the issue of endogeneity bias.

1. Step 1: Choice of the Baseline Model Specification

We answer the question of whether hosts are mimicking hospitality firms by estimating the probability that a host i using Airbnb is also present on k_i alternative rental platforms (*PLATS*), given its type (HT_j) and a set of controls X . Controls include CP_n , PT_m , *SUPER*, *POOL*, *ROOMS*, and *PRICE*. This amounts to estimating a count data model.

The first step was choosing the correct specification between a traditional Poisson regression model and a negative binomial regression model (Cameron & Trivedi, 2013). The Negative Binomial model generalises the Poisson model in the sense that it relaxes the assumption that variance equals mean, allowing for overdispersion in the data. Using the NBREG command in STATA 15 software, we estimated the negative binomial model presented below to test for the presence of overdispersion in the data.

The negative binomial model (Eq. (1)) estimates the probability of observing the presence of a host on a number k of platforms other than Airbnb, conditional on the expected mean parameter λ and the overdispersion parameter $\theta > 0$, as follows:

$$P(PLATS = k) = \left(\frac{\theta}{\theta + \lambda}\right)^\theta \times \frac{\Gamma(\theta + k)}{\Gamma(k + 1)\Gamma(\theta)} \times \left(\frac{\lambda}{\theta + \lambda}\right)^k \quad (1)$$

where Γ is the gamma function defined as

$$\Gamma(t) = \begin{cases} \int_{x=0}^{\infty} x^{t-1} e^{-x} dx & \text{for positive non-integer } t \\ (t-1)! & \text{for positive integer } t \end{cases}$$

The variance of the negative binomial model is given by $v = \lambda + \frac{\lambda^2}{\theta}$.

In summary, in the parameterised version of the model it implements, STATA 15 estimates a parameter $\alpha = \frac{1}{\theta}$. That is, if $\alpha = 0$, there is no overdispersion, and the appropriate model is a traditional Poisson

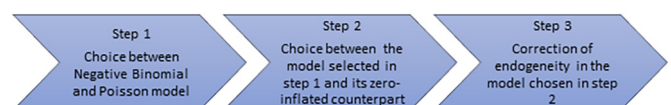


Fig. 1. Three steps of the econometric strategy.

Regression. Our results, detailed in Table 5 of Section 5, show that α is statistically different to 0 at the 1% level ($\hat{\alpha} = 0.190, p \text{ value} = 0.000$). Thus, our baseline specification is a negative binomial regression model.

2. Step 2: Zero-Inflated Models to Manage the Excess of Zeros

Step 1 of our methodology defines our reference model specification. Nonetheless, the data-generation process raises a specific issue. If the underlying assumptions regarding the use of the same picture to promote the property on different platforms and/or the accuracy of the Google Cloud Vision API are not fulfilled, *PLATS* is always 0 for this host. Thus, two groups exist in our sample: a group *A* that contains hosts for whom the assumptions are not satisfied, and a group \bar{A} for hosts satisfying these assumptions.

As Mullahy (1986) demonstrates, in that case, if the probability of being in group *A* for host *i* is p_i and is significantly different from 0, the standard negative binomial or Poisson distributions are no longer appropriate. This problem is known as the excess of zeros issue and is due to the following: the data-generation process does not arise from a single distribution but from two distinct distributions.

Fig. 2 shows that a count of 0 was associated with 49.49% of the hosts. Thus, we had to check whether the excess of zeros was an issue in our study and use an appropriate fix if necessary.

A common fix is to use zero-inflated count models that allow modelling of the excess of zeros and allow for overdispersion. For readability, we present only the basics of the method and what is required to understand the results. See Cameron and Trivedi (2013) for a complete technical presentation.

For each observation in the sample, two possible data-generation processes exist. Process *A* is chosen with probability p_i , and process \bar{A} is chosen with probability $1 - p_i$. Counts k_i generated from process *A* are always zero (so-called degenerated 0); process \bar{A} generates counts that are greater than or equal to 0 from a negative binomial or Poisson distribution.

Process *A* is estimated through a logit model predicting the occurrence of a behaviour that is not in line with our assumptions: the probability that *PLATS* is always equal to 0. Process \bar{A} is estimated using a standard negative binomial or Poisson distribution. The expected count of *PLATS_i* depends on both processes and is obtained as described in Eq. (2)

$$E(PLATS_i) = p_i \times 0 + (1-p_i) \times e^{a_i} \tag{2}$$

where p_i is the probability of being in group *A* obtained from the logit (usually called ‘inflate’) part of the model, and e^{a_i} is the expected count of platforms other than Airbnb and was obtained from the negative binomial or Poisson part of the model.

The zero-inflated model estimates, from a set of independent variables *Z*, demonstrate the probability that *PLATS* always equals 0, and this probability is then used to obtain a correct count distribution. In this paper, we

had to estimate the probability of always observing 0 because some hosts do not use the same picture to advertise a listing on different platforms. We use the variable *PHOTOS* as an explanatory variable because it seems obvious that the size of the set of photos available influences the host’s choice when posting the listing on a platform.

At this stage of the method, we made two decisions. First, we had to replicate step 1, choosing between a zero-inflated negative binomial distribution and a zero-inflated Poisson distribution. Second, we had to decide whether the zero-inflated model improved the negative binomial specification used in step 1.

Once the zero-inflated negative binomial model was estimated using the ZINB STATA command, we could observe that the estimated overdispersion parameter was equal to 0.064 and not significantly different from 0 at the 10% level ($\alpha = 0.064, p - \text{value} = 0.108$). As Mullahy (1986) emphasises, the presence of overdispersion in the data may be due to the excess of zeros. This case was the case here, implying the choice of the zero-inflated Poisson specification over the zero-inflated negative binomial specification. We subsequently had to decide whether the zero-inflated Poisson specification improved our baseline negative binomial model.

Table 4 presents the measures of fit that Long & Freeze (2014) suggest for this type of model. They show that the zero-inflated Poisson specification is more appropriate. Both the maximum difference and the mean difference between predicted and observed count are lower (in absolute terms) in the zero-inflated Poisson specification. Second, this is also true for the sum of absolute differences between predicted and actual probabilities. In addition, the χ^2 fit test establishes that although the negative binomial predicted distribution is significantly different from the actual distribution at the 1% level, the zero-inflated Poisson distribution does not significantly differ from the sample distribution. Last, both the Akaike and Bayesian information criteria are minimised by the zero-inflated Poisson specification.

3. Step 3: The Two-Stage Residual Inclusion Method to Manage Endogeneity

Our explanatory variable of interest is the host type. However, many variables likely to be highly correlated to the host type cannot be observed. For example, the host type is likely to be influenced by the host’s wealth, level of income, age, or ownership of a second home. This implies that our main explanatory variable is potentially correlated to the error term, leading to biased estimates of the coefficients.

To solve this problem, we use an instrumental method called two-stage residual inclusion proposed by Terza, Basu, and Rathouz (2008) and praised by Wooldridge (2014). Geraci, Fabbri, and Monfardini (2016) demonstrate that this method is valid for count data models. In the first stage of the method, auxiliary equations are estimated for each of the variables (or levels of a multinomial variable) considered potentially endogenous. Predicted values are then computed for each of these variables. Residuals are obtained as the difference between observed and predicted values for each variable suspected of endogeneity. The originality of the method appears in the second stage. Potentially endogenous variables are conserved

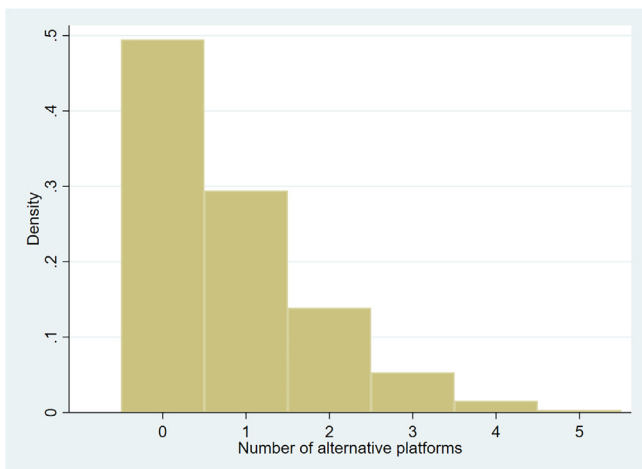


Fig. 2. Observed sample distribution.

Table 4 Comparison of fit between negative binomial and zero-inflated models.

Comparison of mean observed and predicted count		
	Negative binomial	Zero-inflated poisson
Max difference	-0.022	-0.017
Mean difference	0.008	0.005
Comparison of predicted and actual probabilities		
	Negative Binomial	Zero-Inflated Poisson
Sum of absolute differences	0.048	0.028
Pearson’s χ^2	14.646***	5.665
Information criteria		
	Negative Binomial	Zero-Inflated Poisson
Akaike Criterion	9345.071	9326.323
Bayesian Criterion	9432.871	9421.015

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

in the estimation of the equation, and residuals obtained in the first step are simply used in the model as additional explanatory variables.

In our study, we want to check for the exogeneity of the categorical variable HT_j . This implied that in the first stage of the two-stage residual inclusion method, we had to estimate the probability of each host being in a given category. As an auxiliary model, we estimated a multinomial logit in which the explanatory variables, in line with Terza et al., were both the controls used in the main model and a set of four instruments (see Appendix A for details).

Our set of instruments included three qualitative variables (TV , $DRYER$, $CRIB$) and a continuous variable taking only positive values ($BATHS$). The first three instruments are amenities that may or may not be offered by the host. The fourth instrument is the average number of bathrooms on the property. Table 7 presents these variables in detail.

Because the results of the first stage estimation are not central to this paper, they are presented in Table B of Appendix B. We wish simply to underline that the model performed well according to the standard performance of this type of model (Greene & Hensher, 2010). Both the Count R^2 and the Adjusted Count R^2 are good¹ ($CountR^2 = 0.724$ and $Adj. R^2 = 0.160$)².

The result of this multinomial logit estimation is used to compute residuals to be included in the main count model. In accordance with Geraci et al. (2016), we computed standardised residuals. In this context, Geraci et al. show that the power of the endogeneity test is higher when standardised residuals are used rather than raw residuals. We define P_{ij} , a variable with value 1 if the host belongs to host type j and 0 otherwise, as a binary formulation of HT_j . The predicted probability for host i to be of type j is \hat{P}_{ij} . Standardised residuals are then defined as in Eq. (3):

$$\hat{r}_{ij} = \hat{P}_{ij}^{-\frac{1}{2}} (1 - \hat{P}_{ij})^{-\frac{1}{2}} (P_{ij} - \hat{P}_{ij}) \quad (3)$$

The results of the second step of the method are presented in the last column of Table 5 and are discussed in the next section. For now, we highlight that a Wald test on the coefficients associated with the standardised residuals in the main model shows that they are jointly different from 0 at the 5% level ($\chi^2(2) = 8.24$, $pvalue = 0.016$). The variable HT_j is endogenous, and the zero-inflated Poisson model must be augmented to include predicted standardised residuals. Otherwise, the estimated coefficients associated with HT_j are biased. Reliable results are obtained only when we consider the zero-inflated Poisson specification that includes predicted residuals as additional explanatory variables.

5. Results

In this section, we first present general results related to control variables accounting for some characteristics of the hosts. Next, we focus on the influence of host type, defined by the number of listings managed, on multichannel behaviour.

5.1. Impact of some hosts' characteristics

The results of the different econometric models estimated are summarised in Table 5. Columns (1), (2), and (3) present the respective estimations of the negative binomial, zero-inflated Poisson and augmented zero-inflation Poisson (second step of the two-stage residual inclusion method) specifications. Notably, standard errors are robust for all the models estimated in the paper to account for potential specification issues.

Notably, coefficient estimates of models (1) and (2) are similar. We take advantage of this similarity to focus on the interpretation of the estimates of the specifications (2) and (3). However, the similarity of some estimates should not confuse the reader; the relevance of estimating a zero-inflated

model rather than a traditional count model is clear. We eliminate overdispersion and improve the fit of the model (Table 4).

Some coefficients associated with the controls are of interest. The coefficients associated with PT_{APART} and PT_{HOUSE} are strongly significant with a negative sign, indicating that hosts without a typical property type are, *ceteris paribus* (all other things being equal), present on a higher number of platforms. This could, to some extent, be related to individuals with various properties for rent, such as real estate agencies.

The significant and negative coefficient associated with the particular cancellation policy 'moderate', CP_{MODE} , shows that those hosts are present on a lower number of alternative platforms. Notably, the STRICT30 cancellation policy, available by invitation from Airbnb staff to hosts connected to the platform by third party software, is significant in models (1) and (2) but not in model (3). The correction of endogeneity dissipates the effect, suggesting that some unobserved characteristics of the hosts that affect HT_j also affect CP_{MODE} . This could be related the following: this type of cancellation policy is mainly adopted by professional players because of Airbnb's conditions of use.

The variable $POOL$ is positively and significantly associated with $PLATS$. If we consider the coefficient of model (3), hosts who typically offer a pool use alternative platforms 31.26% more than average. This finding suggests that the presence of such a luxury amenity creates an incentive for the host to use several channels. The variable $SUPER$ is strongly and negatively associated with the number of platforms. In model (3), we observed that *ceteris paribus* (all other things being equal), Superhosts use alternative platforms 21.34% less than average. This finding is notable because the Superhost label acknowledges the quality of service provided by the host. That a Superhost uses fewer channels suggests that this type of host develops a different strategy to attract guests (Liang, Schuckert, Law, & Chen, 2017).

5.2. Impact of host type on multichannel practice

Regarding the role of host type on the propensity to use different channels, the model in column (2) highlights clear, strong results. The coefficient associated with HT_2 is 0.227. Thus, *ceteris paribus* (all other things being equal), hosts managing two properties use alternatives to Airbnb 25.5% more than hosts managing one property do. The propensity to use different platforms is even higher for hosts managing more than two listings. With a coefficient estimate of 0.498, they use alternatives to Airbnb 1.65 times more than hosts managing only one listing do. The more properties managed, the more alternative platforms used.

However, the appealing conclusion, in line with the intuition supported by the literature, that hosts managing several listings are professionals is incorrect. Model (3), which accounts for the endogeneity of the host type, demonstrates that the behaviour of hosts managing two listings is no different from the behaviour of hosts managing one listing because the coefficient estimate of -0.166 is not significantly different from 0. Furthermore, the coefficient associated with HT_3 is 0.638 in that case. These changes are due to the significant effect of the predicted residuals on the number of platforms.

The number of alternative platforms used by hosts managing more than two listings is not 65% higher than the number used by type 1 hosts, as suggested by the zero-inflated Poisson model, but 89.3% higher. Endogeneity causes overestimation of the use of alternatives to Airbnb by type 2 hosts and underestimation of the use of alternatives by type 3 hosts.

In specifications (2) and (3), the variable $PHOTOS$ is strongly significant and negatively associated with the probability of observing a degenerate 0. The zero-inflated Poisson model estimates, for each host, a probability p_i of observing a degenerate 0, namely, for this host, we always observe $PLATS = 0$ because the assumptions regarding the picture used to advertise the listing are not fulfilled. The negative sign of the coefficient means that the larger the average number of $PHOTOS$ a host displays, the less likely a degenerate 0 will be observed. The estimate of the average probability p_i in specification (3) is $\hat{p}_i = 12.79\%$ ($p\ value = 0.000$). Our working

¹ $CountR^2 = \frac{\text{Number of correct predictions}}{N}$

² $Adj. CountR^2 = \frac{\text{Number of correct predictions} - N_i}{N - N_i}$ where N_i is the number of observations within the modal class

Table 5
Results from different count model specifications.

PLATS	(1)		(2)		(3)	
	Negative binomial model		Zero-inflated poisson model		Corrected zero-inflated poisson model	
	coef.	p-values	coef.	p-values	coef.	p-values
			Main		Main	
HT ₂	0.223***	0.000	0.227***	0.000	-0.166	0.356
HT ₃	0.527***	0.000	0.498***	0.000	0.638***	0.000
PT _{APART}	-0.161**	0.002	-0.142**	0.005	-0.136*	0.030
PT _{HOUSE}	-0.173***	0.000	-0.172***	0.000	-0.165**	0.004
CP _{FLEX}	-0.113	0.168	-0.071	0.375	-0.092	0.421
CP _{MODE}	-0.396***	0.000	-0.400**	0.0030	-0.409**	0.002
CP _{STRICT}	0.026	0.704	0.024	0.724	-0.006	0.952
CP _{STRICT30}	0.485*	0.021	0.487*	0.012	0.347	0.087
SUPER	-0.216*	0.028	-0.232*	0.019	-0.240*	0.016
POOL	0.294***	0.000	0.264***	0.000	0.272***	0.000
ROOMS	-0.012	0.542	-0.024	0.219	-0.029	0.181
PRICE	0.000	0.513	0.000	0.438	0.000	0.430
ŝ _{HT₂}	-	-	-	-	-0.152*	0.018
ŝ _{HT₃}	-	-	-	-	0.055	0.265
CST	-0.259**	0.003	-0.092	0.298	-0.017	0.922
PHOTOS	-	-	Inflate		Inflate	
CONSTANT	-	-	-0.078*	0.012	-0.084**	0.007
α	0.190***	0.000	-0.731*	0.030	-0.688*	0.040
Akaike	9345.071	-	9326.323	-	9323.435	-
Bayesian	9432.871	-	9421.015	-	9430.073	-
χ ²	303.445***	0.000	270.800***	0.000	286.256***	0.000
N	3916	-	3916	-	3916	-

*p < 0.05, **p 0.01, ***p < 0.001.

assumptions are roughly fulfilled for 87.20% of the sample. The overall estimated probability of observing a 0 count is 48.75% ($\hat{P}(PLATS = 0) = 48.75\%$, p value = 0.000). Hence the probability for a given host to post listings on at least one alternative platform is 51.25%. A majority of hosts practice multichannel distribution.

Table 6 displays the average probability computed from specification (3), for the full sample and per host type, of using a given number of alternative platforms. Probabilities are obtained by using the delta method. This highlights some important findings. Table 6 confirms that the behaviour of hosts managing one listing is not significantly different from the behaviour of those with two listings. In addition, the probability decreases dramatically as the number of platforms increases. The probability of using Airbnb only is 36%, and that of using Airbnb along with another platform is 30.6%. The probability of using two alternatives is 14.1%. Hosts tend, on average,

Table 6
Estimated probabilities for the different HT_j categories.

	Full Sample		HT ₁	
	coef.	5% robust CI	coef.	5% robust CI
Airbnb only	0.360	[0.327;0.392]	0.380	[0.335;0.424]
Airbnb + 1	0.306	[0.293;0.319]	0.307	[0.294;0.319]
Airbnb + 2	0.141	[0.136;0.147]	0.132	[0.118;0.146]
Airbnb + 3	0.048	[0.044;0.051]	0.041	[0.030;0.051]
Airbnb + 4	0.013	[0.012;0.015]	0.010	[0.005;0.015]
Airbnb + 5	0.003	[0.003;0.004]	0.002	[0.001;0.004]
	HT ₂		HT ₃	
	coef.	5% robust CI	coef.	5% robust CI
Airbnb only	0.430	[0.340;0.520]	0.188	[0.121;0.255]
Airbnb + 1	0.297	[0.268;0.325]	0.274	[0.234;0.315]
Airbnb + 2	0.109	[0.073;0.145]	0.213	[0.192;0.233]
Airbnb + 3	0.029	[0.014;0.044]	0.117	[0.082;0.152]
Airbnb + 4	0.006	[0.002;0.010]	0.052	[0.025;0.078]
Airbnb + 5	0.001	[0.000;0.002]	0.02, 0	[0.005;0.034]

to use a limited number of platforms. Nonetheless, the proportion of hosts using Airbnb and at least two other platforms is far from marginal. The cumulative probability of using at least two alternatives to Airbnb is 20.53% (p value = 0.000). A significant proportion of hosts mimic the behaviour of hospitality firms to attract more guests.

Furthermore, the examination of probabilities conditional on the host type in Table 6 is very informative. First, hosts managing many listings are less likely to limit their activity to Airbnb. The probability that they operate only on Airbnb is more than twice as low as that of other types of host. Notably, the probability of operating on Airbnb and one alternative platform is very similar between groups. This finding indicates that using two platforms (Airbnb + 1) is common behaviour, irrespective of host type. Conversely, as the number of alternatives to Airbnb increases, the gap between type 3 hosts, on the one hand, and type 1 and 2 hosts, on the other hand, increases. The probability of using two alternatives is already 1.6 times higher for type 3 hosts. There is a factor of 5 between the probability of type 3 and type 1 hosts respectively using Airbnb and four alternatives. This factor increases to 10 for the use of five alternatives.

This finding is of particular relevance because it confirms that hosts managing many listings are more likely to mimic hospitality professionals. We can easily compute that the probability of a type 3 host using at least Airbnb + 2 alternatives is 40%. But this finding is not the whole story. A total of 18.5% of type 1 hosts who are perceived as very small players because they rent only a single property are also present on at least three platforms. They are mimicking the proactive behaviour of hospitality firms to increase their visibility.

6. Discussion and conclusion

According to our review of the literature, we have presented the first work on multichannel distribution in the short-term rental market. Using appropriate count models, we establish that the parallel use of several

rental platforms is very common behaviour among Airbnb hosts. Notably, more than half of the hosts are likely to practice multichannel distribution. Second, we show that the propensity to use different distribution channels depends, to a significant extent, on the professionalisation level of the host. Hosts with a larger number of properties for rent are more inclined to diversify their distribution channels. Furthermore, even if hosts with one or two properties for rent are, on average, less likely to use several platforms, a significant proportion of them, approximately 20%, are present on Airbnb and at least two other platforms. Overall, the results suggest that professional hosts mimic the behaviour of hospitality firms to enhance their economic performance. Nonetheless, many small hosts, usually perceived as non-professional, are, notably, willing to optimise their occupancy rate through channel diversification. Our work facilitates research related to a new question. We show that hosts establish a portfolio of platforms to improve the performance and/or visibility of their listings.

These findings support important implications for stakeholders and researchers. Recent articles managing hosts' attitude have emphasised the importance of psychological attachment (Lee, Yang, & Koo, 2019) and trust in the platform (Wang, Asaad, & Filieri, 2019). Our results suggest that the attachment to the platform is weak for a large proportion of the host population.

From the platforms' point of view, it strengthens the need to develop adequate strategies to retain hosts. Notably, multichannel distribution means lower profitability for the platform, and the ability to retain hosts is a major managerial issue for Airbnb and its competitors. For example, VRBO is launching its 'Fast Start Program', designed to attract the most efficient hosts from other platforms (VRBO, 2021).

Our contribution also highlights a challenge for policy makers. Before the COVID-19 outbreak, regulations were implemented to limit the adverse effect of short-term rental on local communities. We show that these regulations must be designed at the host level rather than at the platform level. That is, the problem is not to control the activity of Airbnb or Booking but to control the activity of a given host over different platforms. For example, the so-called ELAN law enacted in France in 2019 allows Paris to create a registry of hosts to monitor the short-term rental market, and this practice has been beneficial.

From the hosts' standpoint, we provide insights into an infrequently discussed dimension. Depending on their motivations and goals, they are going to define different platforms portfolios. The theoretical economic literature on two-sided markets has emphasised the possibility of multihoming on web platforms (Armstrong, 2006; Rochet & Tirole, 2003). To mitigate competition between them, hosts have an incentive to simultaneously use several platforms.

Our work also provides useful information and tools to researchers. Ignoring the existence of multiplatform practice could bias the perception of the short-term rental market and affect the reliability of some academic results. The identification of listings posted on competing platforms creates a reliable indicator of the extent of overlapping. Hence, our method provides an enhanced measure of the effective size of the supply side of the short-term rental market. It will help further research to manage the aforementioned bias.

The competing platforms used by Airbnb hosts differ in nature and size. Thus, the major global platforms (e.g. VRBO, TripAdvisor, Booking.com), regional platforms (e.g. Media Lite Group, Locasun Group), and national platforms (e.g. Mediavacances, Gites de France) are competing with each other. The diversity of actors is consistent with the theoretical work of Belleflamme and Peitz (2019). They show that competition between sellers on a given platform can generate differentiation between competing platforms. Differentiation allows the coexistence of both large and popular platforms, in which competition is crucial, and small platforms, on which sellers are protected from competition. This is in line with our empirical results and raises a question for further research. How will the relationships between platforms evolve in a context in which a multichannel strategy is a common behaviour?

Our article is the first attempt and thus has limitations. Similar to many tourist destinations, Corsica is a sea, sun, and sand destination characterised by a short-term rental market for second homes since the 1970s. The picture could be slightly different if urban destinations were studied instead. Second, the listing identification is based only on reverse image recognition. Thus, adding the geographical information and the textual information in the listings to the recognition methodology would improve multiplatform identification. Finally, our study is based on data collected before the COVID-19 outbreak; thus, predicting the evolution of this multichannel practice in the post-pandemic tourism industry is difficult. Notably, the current evolutions seem so salient that a close monitoring of the multichannel strategy is necessary to enhance the understanding of the impact of COVID-19 on the short-term rental market.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Instrumental variables

This Appendix provides descriptive statistics and a correlation matrix relative to the instruments used in step 3 of our methodology.

Table 7
Descriptive statistics for the instruments.

	Mean	Std. Dev.	Min.	Max.
TV	Host typically provides a TV 0.888	0.319	0	1
DRYER	Host typically provides a DRYER 0.174	0.379	0	1
CRIB	Host typically provides a CRIB 0.137	0.343	0	1
BATHS	Average number of bathrooms for this host 1.423	0.872	0	8

Table 8
Correlation matrix for the instruments.

	PLATS	HT ₁	HT ₂	HT ₃	TV	DRYER	CRIB	BATHS
PLATS	1							
HT ₁	-0.180*	1						
HT ₂	0.037*	-0.669*	1					
HT ₃	0.197*	-0.598*	-0.196*	1				
TV	0.0110	0.120*	-0.017	-0.140*	1			
DRYER	-0.014	0.126*	-0.057*	-0.1051*	0.101*	1		
CRIB	0.028	0.084*	-0.025	-0.0830*	0.082*	0.127*	1	
BATHS	-0.012	0.1260*	-0.068*	-0.0928*	0.065*	0.36*	0.167*	1

* $p < 0.05$.

Appendix B. First stage of the two-stage residuals inclusion method to handle endogeneity

This Appendix presents the results of the first stage of the method for the case in which our main explanatory variable is HT_j (Table B). First, cancellation policies have a significant impact on the probability of belonging to a given host category. This is also true for property type. Notably, *POOL* has no impact on the probability of belonging to a given host group, but the impact of the average number of bedrooms is significant. Furthermore, *SUPER* is positively and significantly associated with the probability of belonging to HT_3 . The probability of managing many listings is greater for superhosts. Finally, three out of our four instruments are strongly significant.

Table 9
First-stage regression of the method: multinomial logit on HT_j .

	coeff.	p-values
HT ₂		
CP _{FLEX}	-5.578***	0.000
CP _{MODE}	-5.572***	0.000
CP _{STRICT}	-5.334***	0.000
CP _{STRICT30}	-4.485**	0.007
SUPER	0.122	0.533
POOL	-0.035	0.763
ROOMS	-0.315***	0.000
PT _{APART}	-1.017***	0.000
PT _{HOUSE}	-0.889***	0.000
PRICE	0.000	0.370
TV	-0.305*	0.029
DRYER	-0.327*	0.021
CRIB	-0.122	0.396
BATHS	-0.065	0.524
CST	5.568***	0.000
HT ₃		
CP _{FLEX}	-6.454***	0.000
CP _{MODE}	-7.062***	0.000
CP _{STRICT}	-5.670***	0.000
CP _{STRICT30}	-2.298	0.068
SUPER	0.539*	0.011
POOL	-0.167	0.235
ROOMS	-0.534***	0.000
PT _{APART}	-1.850***	0.000
PT _{HOUSE}	-1.893***	0.000
PRICE	0.000	0.115
TV	-0.854***	0.000
DRYER	-0.639***	0.001
CRIB	-0.584**	0.002
BATHS	-0.003	0.980

(continued on next page)

Table 9 (continued)

	coeff.	p-values
CST	6.950***	0.000
COUNTR ²	0.724	–
ADJ. COUNTR ²	0.160	–
Observations	3916	–

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix C. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.annale.2021.100017>.

References

- Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics*, 37(3), 668–691. <https://doi.org/10.1111/j.1756-2171.2006.tb00037.x>.
- Belleflamme, P., & Peitz, M. (2019). Managing competition on a two-sided platform. *Journal of Economics and Management Strategy*, 28(1), 5–22. <https://doi.org/10.1111/jems.12311>.
- Benítez-Aurioles, B. (2018). Why are flexible booking policies priced negatively? *Tourism Management*, 67, 312–325. <https://doi.org/10.1016/j.tourman.2018.02.008>.
- Beritelli, P., & Schegg, R. (2016). Maximizing online bookings through a multichannel-strategy: Effects of interdependencies and networks. *International Journal of Contemporary Hospitality Management*, 28(1), 68–88. <https://doi.org/10.1108/IJCHM-07-2014-0326>.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press.
- CBRE (2017). *Hosts with multiple units—A key driver of airbnb growth: A comprehensive national review including a spotlight on 13 U.S. markets*. American Hotel & Lodging Association https://www.ahla.com/sites/default/files/CBRE_AirbnbStudy_2017.pdf.
- Dredge, D., Gyimóthy, S., Birkbak, A., Jensen, T. E., & Madsen, A. K. (2016). *The impact of regulatory approaches targeting collaborative economy in the tourism accommodation sector: Barcelona, Berlin, Amsterdam and Paris*.
- Gazzoli, G., Kim, W. G., & Palakurthi, R. (2008). Online distribution strategies and competition: Are the global hotel companies getting it right? *International Journal of Contemporary Hospitality Management*, 20(4), 375–387. <https://doi.org/10.1108/09596110810873499>.
- Geerts, W. (2019, August 15). Short-term rentals: How professionalized should they become? Skift <https://skift.com/2019/08/15/short-term-rentals-how-professionalized-should-they-become/>.
- Geraci, A., Fabbri, D., & Monfardini, C. (2016). Testing exogeneity of multinomial regressors in count data models: Does two-stage residual inclusion work? *Journal of Econometric Methods*, 7(1), 20140019. <https://doi.org/10.1515/jem-2014-0019>.
- Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1), 45–56. <https://doi.org/10.1080/10548408.2017.1308292>.
- Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management*, 30(1), 2–20. <https://doi.org/10.1108/IJCHM-09-2016-0540>.
- Google (2021). Overview of crawling and indexing topics [Documentation]. Google Search Central <https://developers.google.com/search/docs/advanced/crawling/overview?hl=en>.
- Greene, W. H., & Hensher, D. H. (2010). *Modeling ordered choices A primer*. Cambridge University Press.
- Gurran, N., & Phibbs, P. (2017). When Tourists Move In: How Should Urban Planners Respond to Airbnb? *Journal of the American Planning Association*, 83(1), 80–92. <https://doi.org/10.1080/01944363.2016.1249011>.
- Ikkala, T., & Lampinen, A. (2015). Monetizing network hospitality: Hospitality and sociability in the context of Airbnb. *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing - CSCW'15* (pp. 1033–1044). <https://doi.org/10.1145/2675133.2675274>.
- Ioannides, D., Röslmaier, M., & van der Zee, E. (2019). Airbnb as an instigator of ‘tourism bubble’ expansion in Utrecht’s Lombok neighbourhood. *Tourism Geographies*, 21(5), 822–840. <https://doi.org/10.1080/14616688.2018.1454505>.
- Karlsson, L., & Dolnicar, S. (2016). Someone’s been sleeping in my bed. *Annals of Tourism Research*, 58, 159–162. <https://doi.org/10.1016/j.annals.2016.02.006>.
- Ke, Q. (2017). Sharing means renting?: An entire-marketplace analysis of Airbnb. *Proceedings of the 2017 ACM on Web Science Conference - WebSci'17* (pp. 131–139). <https://doi.org/10.1145/3091478.3091504>.
- Kracht, J., & Wang, Y. (2010). Examining the tourism distribution channel: Evolution and transformation. *International Journal of Contemporary Hospitality Management*, 22(5), 736–757. <https://doi.org/10.1108/09596111011053837>.
- Krause, A., & Aschwanden, G. (2020). To Airbnb? factors impacting short-term leasing preference. *Journal of Real Estate Research*, 25. <https://doi.org/10.1080/08965803.2020.1826782>.
- Kwok, L., & Xie, K. L. (2019). Pricing strategies on Airbnb: Are multiunit hosts revenue pros? *International Journal of Hospitality Management*, 82, 252–259. <https://doi.org/10.1016/j.ijhm.2018.09.013>.
- Lampinen, A., Airo, & Cheschire, Coye (2016). *Hosting via Airbnb: Motivations and Financial Assurances in Monetized Network Hospitality*. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 1669–1680.
- Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 26(5), 727–750. <https://doi.org/10.1108/IJCHM-08-2013-0367>.
- Lee, H., Yang, S. -B., & Koo, C. (2019). Exploring the effect of Airbnb hosts’ attachment and psychological ownership in the sharing economy. *Tourism Management*, 70, 284–294. <https://doi.org/10.1016/j.tourman.2018.08.017>.
- Li, J., Moreno, A., & Zhang, D. (2016). Pros vs Joes: Agent pricing behavior in the sharing economy (SSRN Scholarly Paper ID 2708279). Social Science Research Network <https://papers.ssrn.com/abstract=2708279>.
- Liang, S., Schuckert, M., Law, R., & Chen, C. -C. (2017). Be a ‘Superhost’: The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, 60, 454–465. <https://doi.org/10.1016/j.tourman.2017.01.007>.
- Long, J. Scott, & Freeze, Jeremy (2014). *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press.
- Magno, F., Cassia, F., & Ugolini, M. M. (2018). Accommodation prices on Airbnb: Effects of host experience and market demand. *The TQM Journal*. <https://doi.org/10.1108/TQM-12-2017-0164>.
- Marc Simeoni Consulting (2015). *Etude sur l’économie informelle en Corse*. Agence du Tourisme de la Corse, 170 <https://www.corsica-pro.com/telechargement/!/telechargement/2015-etude-sur-leconomie-informelle-en-corse>.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3), 341–365. [https://doi.org/10.1016/0304-4076\(86\)90002-3](https://doi.org/10.1016/0304-4076(86)90002-3).
- Oskam, J., van der Rest, J. -P., & Telkamp, B. (2018). What’s mine is yours—but at what price? Dynamic pricing behavior as an indicator of Airbnb host professionalization. *Journal of Revenue and Pricing Management*, 17(5), 311–328. <https://doi.org/10.1057/s41272-018-00157-3>.
- Padlifter (2021). *Channel and property management software*. Padlifter <https://padlifter.com/listing-category/channel-and-property-management-software>.
- Rochet, J. -C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1, 990–1024. <https://doi.org/10.1162/154247603322493212>.
- Schneiderman, E. T. (2014). *Airbnb in the city*. Office of the New-York State Attorney General <https://ag.ny.gov/pdfs/AIRBNB%20REPORT.pdf>.
- Stangl, B., Inversini, A., & Schegg, R. (2016). Hotels’ dependency on online intermediaries and their chosen distribution channel portfolios: Three country insights. *International Journal of Hospitality Management*, 52, 87–96. <https://doi.org/10.1016/j.ijhm.2015.09.015>.
- Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3), 531–543. <https://doi.org/10.1016/j.jhealeco.2007.09.009>.
- Visser, G., Erasmus, I., & Miller, M. (2017). Airbnb: The Emergence of a New Accommodation Type in Cape Town, South Africa. *Tourism Review International*, 21(2), 151–168. <https://doi.org/10.3727/154427217X14912408849458>.
- VRBO (2021). Vrbo launches fast start program to fuel early success of new hosts. <https://www.expediagroup.com/media/media-details/2021/Vrbo-Launches-Fast-Start-Program-to-Fuel-Early-Success-of-New-Hosts/default.aspx> March 29 2021.
- Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, 62, 120–131. <https://doi.org/10.1016/j.ijhm.2016.12.007>.
- Wang, Y., Asaad, Y., & Filieri, R. (2019). What makes hosts trust Airbnb? Antecedents of hosts’ trust toward Airbnb and its impact on continuance intention. *Journal of Travel Research*. <https://doi.org/10.1177/0047287519855135>.
- Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182(1), 226–234. <https://doi.org/10.1016/j.jeconom.2014.04.020>.

Sauveur Giannoni is an associate professor in economics working in the field of tourism and hospitality.

Daniel Brunstein is a geographer specializing in environmental sciences. He is a research engineer.

Florian Guéniot is a computer scientist involved in the development of softwares for economics and social sciences.

Johan Jouve is a Geographic Information System analyst. All the authors belong to UMR 6240 LISA (Université de Corse).