

Intermediaries as quality assessors: Tour operators in the travel industry ☆

Sofronis Clerides^{a,b,*}, Paris Nearchou^a, Panos Pashardes^a

^a *University of Cyprus, Cyprus*

^b *CEPR, United Kingdom*

Received 24 March 2005; received in revised form 27 October 2006; accepted 7 February 2007

Available online 14 February 2007

Abstract

Intermediaries often arise in order to facilitate trade in markets characterized by asymmetric information. In the travel industry, policymakers have tried to address information asymmetries by providing hotel ratings. We show that those ratings are noisy indicators of quality because of the use of nonuniform standards across countries and limited information available to policymakers. Tour operators use their specialized knowledge to produce better ratings which are useful for comparing hotels within a country, as well as across countries. Intermediation by tour operators provides a better matching of quality with price and leads to a more efficient market outcome.

© 2007 Elsevier B.V. All rights reserved.

JEL classification: L15; L83

Keywords: Intermediaries; Middlemen; Information provision; Tourism and travel industry

☆ We thank the Cyprus Tourism Organization for financial support. Clerides thanks the Yale Economics department for its hospitality and the Cowles Foundation for financial support. The research assistance of Pantelis Mitsi and Anna Theodoulou and the valuable comments of Phillip Leslie and Nadia Soboleva are gratefully acknowledged. Seminar and conference participants at Yale University, European University Institute, Athens University of Economics and Business, the 2002 ASSET Conference, the 2003 International Industrial Organization Conference, the 2003 Conference on Tourism Modelling and Competitiveness and the 2004 EARIE Conference also provided helpful feedback. We are solely responsible for any errors.

* Corresponding author. Department of Economics, University of Cyprus, P.O. Box 20537, CY-1678 Nicosia, Cyprus. Tel.: +357 22892450; fax: +357 22892432.

E-mail addresses: s.clerides@ucy.ac.cy (S. Clerides), nearchou@ucy.ac.cy (P. Nearchou), p.pashardes@ucy.ac.cy (P. Pashardes).

1. Introduction

Imperfect information is an important feature of many markets. Its simplest and most widely-studied form is the asymmetric information case where the seller of an object knows its true value while potential buyers do not. Akerlof (1970) pointed out how this informational asymmetry can hinder the proper functioning of markets and lead to suboptimal levels of exchange. In extreme cases, the buyer's inability to assess quality can kill markets for high quality goods. This gloomy prediction is often overcome in practice through various mechanisms such as the provision of warranties and the buildup of reputations. Additionally, in some markets information asymmetries give rise to a new class of agents whose function is to determine the quality of an object and make that information available to potential buyers. By providing that service the intermediary, or middleman, mitigates the information problem and facilitates trade. Both buyers and sellers can benefit from the presence of an intermediary and are willing to pay for its services. Thus the intermediary has an incentive to invest resources in acquiring the skills that enable it to assess quality. Well-known intermediaries include firms like Moody's and Standard & Poor's, whose rating services play a crucial role in the proper functioning of financial markets. Other agencies rate consumer products or services: *Consumer Reports* evaluates consumer products, *Zagat* rates restaurants and the *American Automobile Association* rates hotels.

Governments are also active information providers. Intervention by public agencies typically comes in the guise of minimum quality standards. For example, educational institutions need to satisfy certain criteria in order to have their degrees recognized; airlines must abide by safety rules; eating establishments must follow hygiene standards. Attainment of a minimum standard is often required in order for an establishment to be allowed to operate. Alternatively, the standard may be optional and firms decide whether to conform to it or not depending on the expected response by consumers. In such markets public and private information providers often coexist and provide complementary services. The task of public agencies is limited to determining whether a product or service meets an objectively defined minimum standard. It is then up to private intermediaries to assess the relative quality of the products and services that pass the minimum standard criterion.

In this paper we study a rare example of a market where both public and private intermediaries provide quality ratings. The travel and tourism industry is characterized by a large degree of asymmetric information. Prospective tourists typically need to transact with agents such as hotels and car rental agencies who are located in faraway destinations, often in foreign countries. The quality of the accommodation and the surrounding locality is difficult to assess without first-hand experience. Recognizing this problem, most countries have established national rating agencies whose task is to assess the quality of hotels in their jurisdiction and rate them according to a five-point star system. At the same time, private intermediaries such as tour operators play a prominent role in this market. Tour operators are firms that sell complete vacation packages. Their mode of operation entails three vital intermediary functions. First, they collect objective information about alternative destinations and make it available on the internet or in brochures suitable for comparison shopping. Second, they provide subjective ratings of the different destinations based on both customer feedback and their own assessment of the facility. Third, they bundle together transportation, accommodation and other services and sell them to consumers as a complete vacation package. Bundling minimizes transaction costs, while pooling together tourists headed to the same destination exploits scale economies.¹

¹ The first two tasks are also performed by other entities, such as travel magazines, but on a much smaller scale.

The fact that tour operators (TOs) sell the products they rate is an interesting feature of this industry. The typical information intermediary is a third party that facilitates a transaction but is not directly involved in it. Retailers might be regarded as a type of seller who also provides information. For example, consider a retailer who sells several brands of consumer electronics. When a consumer walks into the store, the retailer provides information about each product. In addition to technical features and specifications, the retailer may also relay to the prospective buyer his own subjective assessment as well as the experience of other customers who have previously purchased the product. The rise of the internet has formalized this process by allowing consumers to rate products they have bought or merchants they have transacted with. Internet retailers such as amazon.com, auction sites like eBay.com and opinion sites like epinions.com are useful sources of information about products and merchants. Retailers implicitly act as quality certifiers by stocking goods of a certain minimum quality.² Importantly, though, retailers will typically stop short of explicitly rating the products themselves.

Tour operators are essentially retailers who intervene in the market by providing their own ratings. This in itself is an indication that government-provided ratings do not accurately convey quality and fail to adequately address the information problem. Tour operators are in a better position to provide accurate quality ratings for at least two reasons. First, each national rating agency sets its own standards, so national ratings are not comparable across countries. Second, national ratings are typically based on checklists of objective and easily verifiable criteria. Unlike tour operator ratings, they rarely incorporate additional information such as consumer feedback.

Our main goal is to test the conjecture that hotel ratings provided by tour operators are more informative than national ratings. To that end we have collected a unique dataset that includes information on holiday packages from the United Kingdom to several Mediterranean destinations. We observe the price and characteristics of each package as well as two quality ratings for each hotel. One is the “official” rating given by the hotel’s national rating agency; the other is the rating provided by the tour operators themselves. The two ratings coincide in most cases, but for about one third of hotels tour operators give either higher or lower ratings than the national agency. We find that there is substantial variation in the standards used by different national raters. As a result, official ratings are not very useful for cross-country comparisons. Tour operators, on the other hand, are found to apply a common standard for all destinations and produce ratings that are comparable across countries. But we also find evidence that disagreements among raters reflect differences in quality assessment, not just standards. This signifies that tour operators incorporate their superior information into their ratings, making them more accurate indicators of quality than nationally provided ratings. Thus tour operator ratings provide useful information for within-country, as well as for cross-country comparisons.

We review the existing literature in Section 2 and describe the industry and data in Section 3. Our empirical strategy and main results are laid out in Section 4. We close with a summary and some concluding remarks in Section 5.

2. Related literature

A substantial theoretical literature has analyzed the performance of markets when information is incomplete. We focus on the asymmetric information case where consumers cannot precisely

² This point was made by [Marvel and McCafferty \(1984\)](#); they use this insight as an explanation for resale price maintenance. [Chu and Chu \(1994\)](#) build on that argument to show how manufacturers of high quality products can signal their quality by distributing through reputable retailers.

assess a product's quality before purchasing it. A series of papers starting with [Klein and Leffler \(1981\)](#) explores ways in which such market environments can sustain equilibria with a positive price–quality relationship, signifying that price serves as a (correct) signal of quality. The main idea behind these papers is that an equilibrium with both high and low quality producers can only be sustained if high quality producers can set prices above competitive levels and earn a positive stream of profits. Otherwise, they will be tempted to cut quality (and thus cost) and increase short-run profits as unsuspecting consumers will purchase the damaged good at the premium price. The mechanism that delivers this equilibrium varies from paper to paper. In [Klein and Leffler \(1981\)](#) firms add credibility to their quality promises by making sunk investments in nonsalvageable capital (such as advertising). In [Shapiro \(1983\)](#) firms build a reputation by selling high quality products at the low quality price upon entering the market, while in [Allen \(1984\)](#) consumers can infer product quality from the firm's choice of price and quantity.

A different strand of the literature has explored the conditions under which intermediaries (middlemen) emerge. In one branch of the literature ([Rubinstein and Wolinsky, 1987](#); [Fingleton, 1997](#)) intermediaries facilitate exchange by speeding up the time-consuming matching process between buyers and sellers. A second branch examines the scope for intermediation in markets where product quality is unobserved. In [Biglaiser \(1993\)](#) and [Biglaiser and Friedman \(1994\)](#), middlemen arise for two reasons. First, they buy more goods than an ordinary buyer and thus have a greater incentive to invest in skills that enable them to detect the true quality of a good. Second, because they are in the market for a long time they may place a higher value on maintaining a reputation and thus be more likely to report the true quality of a good.

The empirical literature on the value of information has only begun to blossom fairly recently. [Jin and Leslie \(2003\)](#) study the impact of government-provided information in the case of restaurant hygiene report cards. They find that the report card system has led to an improvement in hygiene. In a follow-up paper ([Jin and Leslie, 2006](#)) they argue that an effective reputation mechanism provided incentives for a substantial number of restaurants to maintain high quality standards even before the report card system was implemented. [Dewally and Ederington \(2006\)](#) compare the effectiveness of four different strategies (reputation building, certification, warranties, information disclosure) in resolving seller–buyer information asymmetries in the on-line comic book market. They find that certification by a third party is the strongest signal. Reputation is also important, while warranties and information disclosure are less effective. [Jin and Kato \(in press\)](#) study the on-line market for baseball cards. Sellers can either send their card to professionals to have it graded or they can make their own (unverifiable before purchase) claim about the card's quality. Both graded and ungraded cards are traded on the market. The authors make two interesting observations. First, many high quality cards are sold without being graded even though graded cards are sold at a premium that is much greater than the cost of having a card graded. Second, a surprisingly large number of low quality ungraded cards are sold at premium prices, apparently because buyers take the seller's quality claim as accurate. This example is a good illustration of the extent of the information problem in many markets and the important role of intermediary rating services. Finally, we note the existence of a large literature in finance which tries to evaluate the information content of bond ratings.³

3. Industry and data

Millions of North Americans and northern Europeans flock to southern beach resorts every summer. Geography dictates that Americans congregate in the Caribbean while Europeans prefer

³ For a recent example see [Kliger and Sarig \(2000\)](#).

the Mediterranean. The typical tourist is interested in a comfortable hotel, a sandy beach, lots of sunshine, and plentiful food and drink. Package holidays that include all of the above are designed with this tourist in mind. Tour operators provide centralized information about holiday packages to different destinations. This information has traditionally been distributed in glossy printed brochures, while recently it is becoming increasingly available on the internet. It includes the details of the package (length of stay, services included, price, etc.); hotel characteristics (quality rating, amenities such as swimming pools, restaurants, athletic facilities, cultural activities); and characteristics of the locality (distance to beach, distance to a town, etc.). The large number of available destinations and easy access to centralized information give the consumer a wide range of choices.

3.1. The UK market for package tourism

The UK holiday industry is a mature market that goes back many decades. A long history of consolidation has left the industry with four major players: Thomson Holidays (part of The World of TUI group), Thomas Cook, MyTravel Group (formerly Airtours) and First Choice. In 1996 they held about 56% of the market, while the remainder was split among hundreds of small tour operators, none of which had more than 2% of the market.⁴ The four major operators are vertically integrated both upstream into air transportation (charter airline operation) and downstream into retail distribution (ownership of travel agencies). The trend towards vertical integration, completed in the 1990s, prompted an extensive investigation by the UK Monopolies and Mergers Commission (MMC) regarding possible anti-competitive effects. The investigation concluded that the anti-competitive impact of vertical integration was slight.

In 1999 Airtours and First Choice announced their intention to merge. The merger, which would have made them the largest tour operator in the UK, was blocked by the European Commission.⁵ Contrary to the position taken by the MMC, the European Commission ruled that the relevant market for the evaluation of the merger was the *short-haul* foreign package holiday market. Short-haul holidays include mostly Mediterranean destinations, while destinations such as Thailand, Florida and the Caribbean are considered long-haul. The Commission's ruling suggests that we can focus on one of the two market segments. We chose to concentrate on the short-haul market for a number of reasons. First, the short-haul market is more concentrated into the hands of the four major operators. Second, the relative homogeneity of short-haul destinations facilitates cross-country comparisons. And third, the large number of destinations per country gives us enough degrees of freedom to support inference at the country level. In 1998 the four major tour operators held a combined 85.5% of the UK short-haul foreign package holiday market: Thomson had 30.7%; Thomas Cook, 20.4%, Airtours 19.4% and First Choice 15.0%.⁶

3.2. Data

Our data were taken from the 2003 catalogues of the two biggest tour operators, Thomson and Thomas Cook. Each operator publishes a large catalogue (with an enticing name like "Summer Sun") that lists all packages available to various destinations. Several different packages are available for each hotel. The traveler can choose the number of days he will be

⁴ Source: Monopolies and Mergers Commission (1997).

⁵ The decision was later overturned by the European Court of First Instance, but the merger was abandoned anyway.

⁶ Source: AC Nielsen data reported in European Commission (1999).

Table 1
Distribution of hotels by country and star rating

Country	2-star	3-star	4-star	5-star	Total
Balearics	9	62	34	2	107
Canaries		12	75	1	88
Cyprus	1	15	37	5	58
Greece	6	24	55	12	97
Italy		9	48	2	59
Malta		2	13	3	18
Spain	2	44	63	1	110
Tunisia	1	8	21	17	47
Turkey		5	18	3	26
Total	19	181	364	46	610
Of which:					
Thomson	15	119	200	23	357
Thomas Cook	4	62	164	23	253

staying (usually 7, 10 or 14) and the meal option he prefers (self-catering, bed and breakfast, half-board, full-board). For each hotel-year we focus on the seven-day, half-board package in high and low season.⁷ Whenever a half-board package was not available we took the price of an alternative package, usually the full-board one. The data for each package include the price, official star rating, dates and duration of stay, and several hotel and locality characteristics. In addition, Thomson reports satisfaction scores that it collects from its customers. Four ratings are available for each hotel: one for accommodation quality, one for food quality, one for location, and overall satisfaction.

Data collection occurred in two phases. At a first stage we collected complete information (including all characteristics) on 380 hotels, 229 of them from Thomson and 151 from Thomas Cook. This dataset was sufficient for the analysis of Sections 4.1–4.3. At a later stage we added information on another 230 hotels. For this second set of hotels we collected only information required for the analysis in Section 4.4. Descriptive statistics in this section use the full dataset (610 hotels) where possible and the partial dataset (380 hotels) otherwise.

Hotels are located in the following Mediterranean regions: Cyprus, Greece, Italy, Malta, Spain (divided into Mainland, the Balearic Islands and Canary Islands), Tunisia and Turkey. Table 1 shows the distribution of hotels by official star rating in each region.⁸ Four-star hotels are the most popular with both agents, but Thomson has more three- and two-stars hotels in its portfolio. Fifty-four hotels are included in both catalogues. The overlap between the two catalogues is surprisingly limited. It might indicate that agents specialize in different niches of the market, or that hotels prefer to deal with only one operator.⁹

⁷ We chose the second week of August for our high season package and the second week in May for the low season package. We carried out our analysis on high and low season prices separately and got very similar results. Consequently, in the paper we only report results using high season prices.

⁸ Greece presents a problem because it does not rate its hotels according to the usual star system. Instead, it has categories called Deluxe, A, B, and C. We translated this into stars by assigning five-stars to Deluxe hotels, four-stars to category A hotels, three-stars to category B hotels and two-stars to category C hotels.

⁹ The bargaining process between hotels and TOs raises many interesting issues that go beyond the scope of this paper. Our analysis is conditional on the existence of agreements between TOs and selected hotels.

3.3. Hotel quality ratings

Most countries, especially those that try to promote themselves as tourist destinations, have established agencies whose role it is to rate hotels according to a star system. National ratings are typically based on objective and easily verifiable criteria. Essentially, the rating agency compiles sets of features that are necessary for each star category. Such features may include the number of restaurants and swimming pools, availability of sports facilities (tennis courts, volleyball courts, gym, aerobics instructors, etc.), number of wait and cleaning staff, whether rooms have private bath (as opposed to shower), satellite TV, etc.

The obvious shortcoming of national ratings is that they do not lend themselves to cross-country comparisons. Every country's national rating agency has its own set of criteria for awarding stars, hence four-star hotels in one country may on average be of higher or lower quality than four-star hotels in a different country. There are other reasons to question the usefulness of national ratings. Aspects of quality that are not easily quantifiable are often left out of the process in order to avoid controversy. Moreover, it is easy to imagine that indifference and corruption among the civil servants that carry out the inspections could taint the ratings (at least in some countries). Finally, even if the ratings are correct when the stars are first awarded, it is not clear how meticulous national authorities are about revisiting hotel establishments and reassessing their star-worthiness.

Tour operators are well aware of these limitations and many of them have devised their own rating system. The TO's rating of each hotel is provided in the brochures, alongside the official rating. Thus the consumer is able to use the tour operator's rating as an objective indicator of quality that is comparable across countries. The agent's rating is an extremely useful benchmark that we will make extensive use of in this paper. Operators' ratings are based on objective criteria similar to those used by national rating agencies but also – critically – on information they collect from past customers and their own experience with the establishment. Given their improved information content, these ratings should reflect quality better than those provided by national rating agencies.

On the other hand, it is possible that operators are not objective in their ratings. For example, they may systematically upgrade hotels in order to justify higher prices. One could also imagine scenarios where tour operators use the threat of downgrading a hotel as a bargaining tool in price negotiations. From a theoretical standpoint, it is hard to imagine that operators can give misleading ratings on a systematic basis. Such behavior may successfully mislead consumers in the short-run, but it is unlikely to be sustainable in the long run. This is especially true in the case of some operators (such as Thomson), who publish separate customer feedback scores in addition to their own rating. This allows readers to separate out the impact of the operator's assessment of the facility on the overall rating. Such transparency makes it difficult for the operator to manipulate ratings. Indeed, the ensuing analysis reveals no evidence of systematic bias in operator

Table 2
Summary statistics for customer satisfaction ratings

Rated item	Mean	Standard deviation	Minimum	Maximum	<i>N</i>
Location	0.889	0.108	0.365	1	183
Accommodation	0.918	0.075	0.595	1	183
Food	0.819	0.135	0.311	1	172
Overall holiday satisfaction	0.966	0.029	0.811	1	183

ratings. Summary statistics for consumer satisfaction scores are provided in Table 2. Such scores are provided for 183 out of the 229 hotels in our initial sample (172 for food). Note that although the overall holiday satisfaction score exhibits little variation, the other three indices vary by a fair amount, something that will be useful in the econometric analysis.

Table 3 displays a frequency cross-tabulation of official versus operator ratings separately for each operator. The number in each cell is a count of hotels with the corresponding ratings. The cross-diagonal (in boldface) indicates agreement between national and agent ratings. Although the majority of ratings coincide, a substantial number of hotels are off the diagonal, meaning that they get a different rating from the agent than they do from their national agency. Moreover, discrepancies go in both directions: some hotels are upgraded by the agent while others are downgraded. Three-star hotels are more likely to be upgraded by both agents. Thomson is also more likely to upgrade four-star hotels, while Thomas Cook is roughly equally likely to upgrade or downgrade them. Note also that Thomas Cook creates intermediate categories by giving out pluses (“+”).

Thomson is much more likely to upgrade hotels rather than to downgrade them. One interpretation may be that the operator has low standards and upgrades hotels in order to justify higher prices. An alternative interpretation is that in selecting hotels to include in its catalogue Thomson targets undervalued destinations which it then goes on to upgrade. The latter interpretation is consistent with Thomson being the more downmarket operator; more evidence in support of that interpretation will be provided later on.

The fact that 54 hotels appear in both catalogues gives us the opportunity to check whether the two operators' ratings are consistent with each other. A summary comparison is provided in Table 4. Sixty-one percent of hotels (33 out of 54) maintain their official rating in both catalogues and an additional ten hotels were either upgraded or downgraded by both TOs. This leaves eleven hotels (20%) for which the actions of the TOs differ. Six hotels were upgraded by Thomas Cook but not by Thomson, while four hotels received the reverse treatment. The one remaining hotel was downgraded by Thomson but not by Thomas Cook. Reassuringly, there is no case of a hotel that was upgraded by one agent but downgraded by the other. The simple correlation between the ratings of the two operators is .799. By comparison, the correlation between each operator's rating

Table 3
National versus operator ratings

		National rating				
		2-star	3-star	4-star	5-star	Total
Thomson's rating	2-star	6	11	1		18
	3-star	9	79	19		107
	4-star		29	145	5	179
	5-star			35	18	53
	Total	15	119	200	23	357
Thomas Cook's rating	2+ -star	2	1			3
	3-star	2	36	6		44
	3+ -star		24	16		40
	4-star		1	114	9	124
	4+ -star			24	5	29
	5-star			4	9	13
Total		4	62	164	23	253

Table 4
Rating differences among tour operators

Thomson	Thomas Cook			Total
	Downgrade	Same	Upgrade	
Downgrade	2	1		3
Same		33	6	39
Upgrade		4	8	12
Total	2	38	14	54

and the one by the national authority if .609 for Thomson and .712 for Thomas Cook. The fact that the TO ratings are more closely correlated with each other than with the national ratings can be seen as evidence that they assess quality in similar ways. Some of the remaining differences can be attributed to the fact that Thomson takes into account customer feedback in formulating its ratings, while Thomas Cook bases its ratings on “the views of its own managers in the UK and overseas”.

4. Empirical analysis

4.1. Conceptual framework

We begin by sketching a conceptual framework that will guide and motivate our empirical analysis. The relevant market is package tourism; a product in this market is a vacation package that includes – at the very least – accommodation at and transportation to a pre-specified hotel. In addition the package may include meals, excursions, and other services. Products are inherently differentiated and consumers will choose at most one of the numerous options available to them. Hence a discrete choice representation is a natural way to conceptualize this market. Product attributes can be categorized as general, national (or regional) and hotel-specific. General attributes include contract terms that are not specific to the destination, such as the length of stay, transportation to and from the departing airport, time of departure, meals included, level of service. National characteristics include weather, cleanliness, safety, interesting sites, cultural activities, and quality of other local services such as restaurants. Hotel attributes include the various services and amenities being offered and the hotel’s official star rating. Because we are interested in differences across destinations, we abstract from general characteristics by considering a package of fixed contract terms. We denote hotel-specific characteristics by H and national characteristics by N .

National rating agencies observe a subset $H^n \subset H$ of hotel characteristics. Each country n has its own rating system, which is a function $S^n(H^n)$ that maps hotel characteristics onto a star rating $S_n \in \{2, 3, 4, 5\}$. Tour operators observe a set of hotel characteristics $H^o \subset H$, a set of national characteristics $N^o \subset N$, and the national rating S_n . Some operators may also observe a customer satisfaction score for each hotel, denoted by CS. Each operator has his own rating system $S^o(H^o, CS^o, S_n)$. There is no n superscript because the operator applies uniform criteria across all countries. Moreover, national characteristics do not enter as an argument in the rating function because a rating system is strictly an assessment of the hotel. We impose no a priori restrictions on the relationship between H^n and H^o . It is possible that operators observe some characteristics that national agencies do not, and vice versa.

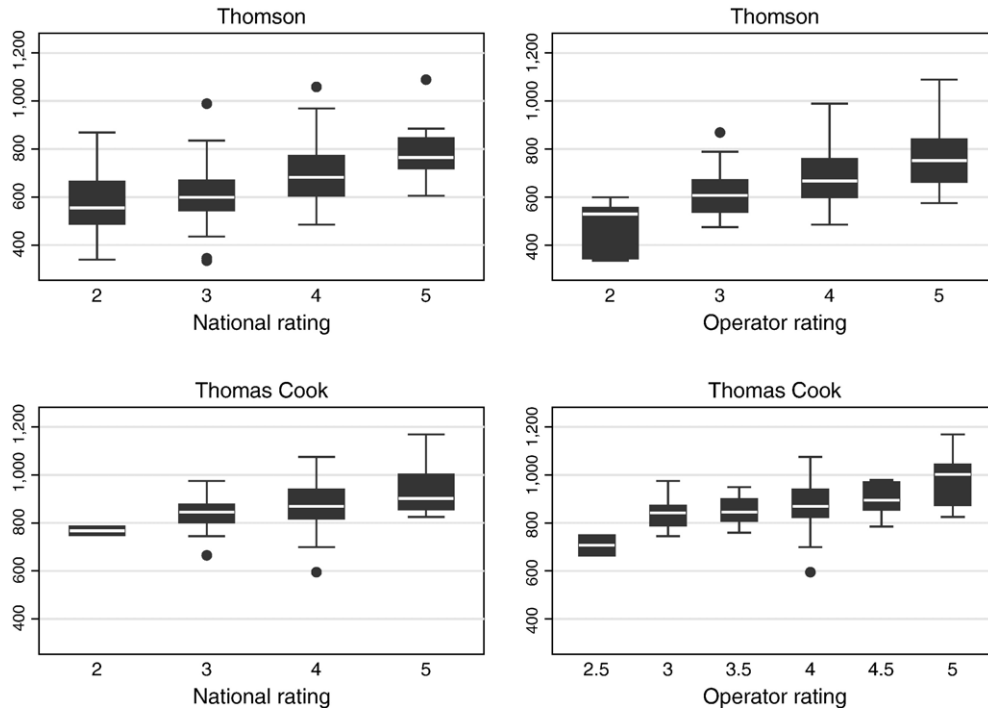


Fig. 1. Distribution of package prices by tour operator and hotel category.

We think of consumers as having preferences over characteristics H and N and over the star ratings S_n and S_o . Consumers value the ratings because they can not observe all elements in H , so the ratings may incorporate additional information. Tour operators take consumer preferences into account in order to formulate a pricing function $P^o(H^o, N^o, S_n, S_o)$. Consequently, the price of a package depends on the characteristics of the region as well as the quality of the individual hotel. The assumption that operators first choose a rating and then a price is a reflection of our claim that operators do not manipulate ratings, something that we will return to later on.

4.2. Pricing functions

As a first step in our empirical analysis we explore the relationship between ratings and prices. We must first note an important distinction between the two catalogues. In addition to the standard services, each operator offers a number of additional options such as allowance for extra luggage, transportation to and from the departing airport, better check-in times, in-flight service, and so on. In addition, the price travelers have to pay depends on their flight's departure time; day flights are more expensive. Our two operators take different approaches to pricing those options. Thomas Cook offers a menu of four preset packages (extra, standard, economy and basic) that travelers can choose from. The price listed in the catalogue is the price of the extra package, and discounts for lesser packages are reported in footnotes. Thomson, on the other hand, allows its customers to choose any combination of options they like. The catalogue reports the price of the basic package and the price of each option. One might interpret this as an indication that Thomas Cook tries to

steer its customers towards the high quality package, while Thomson takes a less targeted approach.

Since quality ratings are discrete, each category covers a range of quality levels. Hence, two holiday packages allocated the same star rating can differ substantially in terms of price. This phenomenon is not observed only between destinations but also within a particular region, where the two packages are rated by the same authority. The distribution of prices by agent and hotel category is summarized in the box-and-whiskers plots in Fig. 1.¹⁰ For each operator we plot the distribution of prices by both national and operator rating. There is considerable within-agent overlap between categories, particularly in the case of Thomas Cook. This overlap is reduced somewhat when hotels are classified by operator rating instead of national rating. But a notable degree of overlap remains nonetheless.¹¹

Pricing patterns of holiday packages can be explored in more detail using hedonic regression. We specify the empirical analogue of the operator's pricing function $\mathcal{P}^o(H^o, N^o, S_n, S_o)$ as:

$$\ln P_i = \alpha + S_i^n \gamma_n + S_i^o \gamma_o + H_i \gamma_h + D_i \gamma_d + \varepsilon_i. \quad (1)$$

S_i^n and S_i^o are vectors of dummy variables representing hotel i 's national and operator rating respectively. The set of country dummies D_i captures national characteristics N . Our aim is to use Eq. (1) to answer two questions. First, are national or operator ratings better determinants of price? Second, how important are country effects in the determination of price? We perform our analysis separately for each agent and present the results in Table 5. We present results from different specifications, suppressing coefficients that are not directly relevant. In column 1 we estimate Eq. (1) without the operator's rating (S_i^o). National ratings are statistically significant and take on reasonable values.¹² In column 2 we add the operator's rating to the regression (that is, we estimate the full Eq. (1)). This addition dampens the coefficients on national ratings substantially, rendering them statistically insignificant. Their explanatory power seems to have been taken over by operator ratings, which take on reasonable size and statistical significance. This indicates that agent ratings are more tightly correlated with prices than official national ratings. As a further test, in column 3 we present the results from estimating the following variant of Eq. (1):

$$\ln P_i = \alpha + \gamma_s \cdot \text{sgn}[S_i^n - S_i^o] + S_i^o \gamma_o + H_i \gamma_h + D_i \gamma_d + \varepsilon_i. \quad (2)$$

The function $\text{sgn}[S_i^n - S_i^o]$ takes the value of 1 if hotel i is upgraded by the operator (or underrated by the national agency), -1 if the hotel is downgraded (overrated) and 0 if the two ratings are the same.¹³ In the case of Thomson, upgraded hotels command a 7.9% premium over hotels of the same national category that are "correctly" rated, while downgraded hotels are 17.7% cheaper. Similar results but on a smaller scale are observed in the case of Thomas Cook: upgraded hotels are 2.8% more expensive while downgraded hotels are 5.1% cheaper.

Country dummies have a lot of explanatory power in the regression, even after we control for all available information. Coefficients on country dummies represent percentage price differences

¹⁰ The "box" indicates the inter-quartile range (from the 25th to the 75th percentile) and the "whiskers" indicate the whole range. The horizontal line in the box denotes the median. Notable outliers are plotted individually.

¹¹ The large difference in the absolute level of prices across operators reflects the fact that for Thomas Cook we have the price of the extra (high quality) package, while for Thomson we have the price of the basic package.

¹² Interestingly, Thomas Cook features less price variation than Thomson.

¹³ Eq. (2) is slightly less general than Eq. (1) in that it restricts the difference $\gamma_n - \gamma_o \equiv \gamma_s$ to be the same for all star categories. Note that in Eq. (1) S_i^n denotes a vector of dummies while in Eq. (2) it denotes a single variable containing the actual rating.

Table 5
Estimates of pricing functions

	Thomson			Thomas Cook		
	(1)	(2)	(3)	(1')	(2')	(3')
National 2-star	−0.091 (0.070)	−0.031 (0.065)	−0.157* (0.066)	−0.122** (0.033)	−0.126** (0.035)	−0.143** (0.033)
National 4-star	0.088** (0.023)	0.015 (0.023)	0.109** (0.021)	0.033* (0.016)	0.031 (0.020)	0.052** (0.018)
National 5-star	0.196** (0.039)	0.074† (0.039)	0.267** (0.037)	0.071** (0.025)	0.045† (0.026)	0.122** (0.031)
Operator 2-star		−0.278** (0.060)				
Operator 4-star		0.065** (0.023)			0.003 (0.023)	
Operator 5-star		0.159** (0.031)			0.081* (0.035)	
Half star					0.001 (0.022)	
Upgraded			0.079** (0.020)			0.028 (0.017)
Downgraded			−0.177** (0.035)			−0.051** (0.017)
Balearics	0.066† (0.038)	0.038 (0.036)	0.045 (0.035)	0.111** (0.018)	0.112** (0.018)	0.113** (0.018)
Canaries	0.041 (0.039)	0.031 (0.040)	0.033 (0.039)	0.080** (0.020)	0.078** (0.020)	0.084** (0.020)
Cyprus	0.119* (0.048)	0.092* (0.042)	0.112* (0.045)	0.115** (0.037)	0.121** (0.039)	0.107** (0.036)
Greece	0.049 (0.040)	0.040 (0.037)	0.036 (0.036)	0.065 (0.026)	0.068 (0.027)	0.062 (0.026)
Italy	0.252** (0.054)	0.257** (0.047)	0.254** (0.047)	0.054 (0.035)	0.050 (0.036)	0.046 (0.036)
Malta	−0.087* (0.043)	−0.094* (0.043)	−0.089* (0.043)	0.073 (0.047)	0.089 (0.054)	0.093* (0.041)
Tunisia	−0.129* (0.052)	−0.136** (0.050)	−0.137** (0.050)	0.031 (0.025)	0.031 (0.026)	0.039† (0.022)
Turkey	0.071 (0.053)	0.032 (0.046)	0.059 (0.047)	0.103** (0.037)	0.086* (0.036)	0.094** (0.035)
<i>N</i>	229	229	229	151	151	151
<i>R</i> ²	0.516	0.631	0.610	0.535	0.553	0.575
<i>F</i> _(<i>N</i>−<i>K</i>,<i>N</i>−1)	10.39	13.29	13.10	6.41	6.18	6.41

Significance levels: †: 10%, *: 5%, **: 1%. Robust standard errors are reported. In addition to reported variables, all regressions include hotel characteristics, half- and full-board identifiers and intercept.

between hotels of similar quality in different countries. The estimates show substantial price variation across countries, but also across agents. In Thomson's catalogue, Italian hotels are substantially more expensive than those of any other country. In Thomas Cook, on the other hand, Italian hotels do not command such a premium. Cyprus, Turkey, Greece and the Spanish island groups enjoy a premium over the Spanish mainland from both agents.¹⁴ Mainland Spain is in fact

¹⁴ We remind the reader that the case of Greece should be interpreted with some caution because of its unique rating system; see footnote 8.

the cheapest region in the Thomas Cook catalogue, while Malta and Tunisia are cheaper in the Thomson catalogue.

The magnitude of the estimated country premia is substantial and in some cases striking. They are likely picking up quality attributes that we do not account for and are shared by all packages to a specific destination. For example, they will pick up any systematic differences in hotel quality that are not corrected by operator ratings. More likely, country premia represent national characteristics (tourism production inputs) that make the vacationer's stay more pleasurable and thus increase the quality of the service being offered. These characteristics could include proximity, weather, auxiliary attractions and services such as restaurants, quality of the infrastructure, cleanliness, and safety. Some destinations may also have linguistic, cultural, or historical ties to different origin countries.¹⁵ It is also possible that some countries may have higher prices because of high domestic demand for tourism; Italy may be an example of that.

The finding that prices are better correlated with operator ratings than with national ratings is hardly surprising; prices are set by operators after all. Nonetheless, the strength of the correlation is noteworthy. Consumer theory gives us the conditions under which our price regressions can be afforded a hedonic interpretation.¹⁶ If those (admittedly quite restrictive) conditions hold, then the estimated coefficients measure the welfare impact of each product characteristic. Under that interpretation, the availability of operator ratings results in a substantial welfare gain.

4.3. National vs operator ratings

Our next task is to explore the rating determination process by estimating empirical analogues of the rating functions $S^n(H^n)$ and $S^o(H^o, CS^o, S_n)$. Let S_i^n and S_i^o be the national and operator ratings of hotel i whose characteristics are described in vector H_i . The hotel's location (region) is encoded in the vector of dummies D_i and customer satisfaction scores are included in vector CS_i . The general form of the equations to be estimated is:

$$S_i^n = \alpha^n + H_i'\beta_h^n + D_i'\beta_d^n + CS_i'\beta_{CS}^n + \varepsilon_i^n \quad (3)$$

$$S_i^o = \alpha^o + H_i'\beta_h^o + D_i'\beta_d^o + CS_i'\beta_{CS}^o + \beta_s S_i^n + \varepsilon_i^o \quad (4)$$

The equations were estimated using an ordered logit specification as the dependent variables are discrete. In columns 1–3 of Table 6 we present the results of estimating Eqs. (3) and (4) with just hotel characteristics and country dummies (omitting satisfaction scores and the official rating). Hotel characteristics turn out to be important determinants of ratings. At the bottom of the table we report the results of testing the null hypothesis that all coefficients on characteristics are zero ($H_0: \beta_h = 0$). The hypothesis is easily rejected in all three cases. Estimated coefficients are qualitatively similar across the three raters. Amenities such as a full bathroom (as opposed to a

¹⁵ For example, Cyprus and Malta were British colonies and maintain close ties with their former rulers.

¹⁶ We write the i th good (holiday package) purchased by the consumer as $q_i^* = \theta_i q_i$, where q_i is the quantity (e.g. 1 week bed and breakfast for one person) and $\theta_i \geq 1$ an index of quality defined by vector of quality characteristics z , i.e. $\theta_i = f_i(z)$, where f_i is increasing in z . From duality we know that maximizing $U = u[q_1^*, q_2^*, \dots, q_n^*]$, subject to $\sum_i q_i^* p_i$, where p_i is the price of the i th good, is equivalent to minimizing the cost $C = \sum_i q_i p_i^*$ of achieving a given U , where $p_i^* = p_i / \theta_i$. Also, if U is increasing then C is decreasing in θ_i . Therefore, defining $f_i(z) = \prod_k z_k^{b_k}$, we arrive at the hedonic regression $\ln p_i = \alpha_i + \sum_k b_k z_k$, where $\alpha_i = \ln p_i^*$.

Table 6
Ordered logit estimates of rating functions

Variable	National	Thomson	ThCook	ThCook	Thomson	National
	(1)	(2)	(3)	(4)	(5)	(6)
Bathroom	2.402** (0.426)	2.673** (0.499)	2.868** (0.890)	1.576 (0.978)	2.240** (0.851)	1.408* (0.584)
Heated pool	1.086** (0.405)	0.814 (0.534)	1.484** (0.539)	0.987 (0.686)	0.827 (0.627)	0.899 (0.703)
Waiter service	0.738* (0.336)	1.386** (0.521)	1.013 (0.712)	1.161† (0.659)	0.671 (0.535)	0.808† (0.481)
Video games	−0.798* (0.372)	−0.042 (0.459)	−0.775† (0.450)	−0.011 (0.412)	0.256 (0.620)	−0.062 (0.623)
Mini golf	0.059 (0.365)	0.855† (0.454)	0.259 (0.465)	0.297 (0.471)	1.203† (0.643)	0.154 (0.568)
Aerobics	0.839* (0.360)	0.313 (0.340)	0.261 (0.420)	−0.001 (0.408)	−0.633 (0.586)	1.006 (0.739)
Coaching	−0.548 (0.453)	−0.802† (0.484)	−0.452 (0.500)	0.245 (0.605)	0.078 (0.831)	−1.285 (1.229)
Gym	1.503** (0.330)	1.186** (0.345)	1.130** (0.419)	0.607 (0.445)	0.010 (0.470)	1.393* (0.561)
Beauty salon	1.088** (0.377)	1.819** (0.620)	1.450* (0.564)	1.338† (0.685)	1.214 (0.783)	1.448** (0.543)
Satellite TV	1.217** (0.322)	1.486** (0.385)	0.747 (0.488)	0.285 (0.535)	0.938* (0.455)	1.369* (0.581)
Balearics	−0.879* (0.389)	0.472 (0.428)	−0.980† (0.568)	−0.199 (0.674)	−0.022 (0.826)	−1.482* (0.693)
Canaries	0.988* (0.441)	1.078† (0.642)	0.095 (0.606)	−0.151 (0.739)	0.230 (0.886)	0.939 (0.637)
Cyprus	−0.258 (0.638)	0.026 (0.569)	1.912* (0.829)	1.399 (0.862)	1.569 (0.990)	−1.109 (0.975)
Greece	2.484** (0.479)	1.803** (0.515)	1.608* (0.650)	0.28 (0.636)	−0.123 (0.996)	2.233** (0.791)
Italy	2.916** (0.531)	2.303** (0.878)	1.996** (0.757)	0.585 (0.837)	0.352 (1.211)	3.105** (0.856)
Malta	1.881** (0.728)	1.016 (0.889)	−0.105 (0.712)	−1.454† (0.865)	−0.994 (0.993)	2.010* (0.968)
Tunisia	1.331* (0.740)	0.479 (0.792)	−0.185 (0.944)	−0.871 (1.055)	0.678 (1.059)	1.371 (1.129)
Turkey	2.834** (0.660)	1.921** (0.591)	2.872* (1.280)	1.17 (1.429)	−1.339 (1.319)	2.966** (1.132)
Official rating				3.451* (0.440)	4.043** (0.627)	
Customer satisfaction scores for:						
Accommodation					12.07* (5.992)	10.08** (3.750)
Food					6.637** (1.964)	2.588† (1.573)
Location					−1.237 (2.567)	4.614* (2.108)
Overall					−12.44 (12.68)	−22.29* (10.74)
N	362	229	151	151	172	172

(continued on next page)

Table 6 (continued)

Variable	National	Thomson	ThCook	ThCook	Thomson	National
	(1)	(2)	(3)	(4)	(5)	(6)
Log-likelihood	−247.9	−196.9	−165.0	−129.5	−94.01	−114.8
χ^2 tests (p -values in parentheses):						
H_0 : All coefficients=0	148.6 (.000)	108.0 (.000)	60.8 (.000)	126.0 (.000)	89.95 (.000)	89.00 (.000)
H_0 : $\beta_h=0$	102.3 (.000)	91.54 (.000)	46.02 (.000)	14.08 (.170)	14.45 (.153)	29.87 (.001)
H_0 : $\beta_d=0$	74.13 (.000)	20.54 (.009)	28.22 (.000)	11.31 (.185)	8.20 (.415)	39.39 (.000)

Robust standard errors are reported. Significance levels: †: 10%, *: 5%, **: 1%.

shower), heated pool, waiter service, gym, beauty salon and satellite TV count towards a higher rating. The negative coefficient on video games is not surprising because they are often a feature of low-end establishments.

Coefficients on country dummies can tell us whether operator ratings are successful in making their ratings comparable across countries. We know that country effects should be important in the national rating function in (3) because each country has its own rating system. We therefore expect that country dummy coefficients (β_d^c) will be statistically different from zero. On the other hand, country effects should not – in principle – affect tour operator ratings; we should have $\beta_d^o=0$. But this will only be true if we have the same information as tour operators; that is, if $H_i=H^o$. If there are unobserved elements in H that are correlated within-countries, this will show up in our regression as country effects ($\beta_d^o \neq 0$).

As expected, most country dummies enter significantly in the national rating regression. Some of them are also statistically different from zero in the operator regression, but the coefficients in the operator regression are generally smaller in magnitude than the ones in the national regression and contribute less to the likelihood function. In the last row of the table we present test statistics for the null hypothesis that the country dummies are all zero (H_0 : $\beta_d=0$). The null is rejected for both regressions, but more strongly so in the national regression. The fact that country dummies remain significant in the operator regression suggests either that a country bias persists in operator ratings or that we did not adequately control for quality. The latter is more likely because we have not included any information that is private to tour operators, namely customer satisfaction scores. We do so in the full specification presented in columns 4–6. We have no customer satisfaction scores for Thomas Cook hotels so we add only one variable, the national rating.¹⁷ Doing so is enough to soak up almost all statistical significance from the characteristics and country dummies. Only two characteristics and one country dummy remain significant at anything better than the 10% level. The hypothesis that all coefficients in each group are zero (H_0 : $\beta_h=0$ and H_0 : $\beta_d=0$) cannot be rejected. In the case of Thomson hotels the full specification includes customer satisfaction scores.¹⁸ The results reported in column 5 are reassuring. First, as in the Thomas Cook case, country dummy coefficients become statistically insignificant. Second, the impact of customer satisfaction scores is consistent with conventional wisdom and with some of our earlier assumptions. A hotel's rating depends on the quality of accommodation and food, as it should. It does not depend on location and on the customer's overall satisfaction with his vacation, as it should not.

¹⁷ An alternative specification would involve moving the national rating to the left-hand side and using the difference between the ratings as the dependent variable. We implemented this and obtained very similar results.

¹⁸ This reduces our sample to 172 observations.

Table 7
Tests for equality of coefficients across regressions

Test of equality of coefficients on:	TUI vs National	TC vs National	TUI vs TC
Hotel characteristics (β_h , $df=10$)	10.92 (0.3642)	5.08 (0.8856)	5.36 (0.8660)
Country dummies (β_d , $df=8$)	28.23 (0.0004)	22.11 (0.0047)	16.21 (0.0394)

Note: TUI = Thomson, TC = Thomas Cook. *P*-values are in parentheses.

In column 6 we present the results from adding Thomson's customer satisfaction scores to the national rating function. The rating agency does not have access to the scores, but they may be correlated with some hotel attributes that the agency observes and we do not. Indeed, the satisfaction score for accommodation comes out strongly significant, suggesting that national rating agencies do take into account some of the things that make consumers happy. Food is more difficult to assess and consequently does not seem to matter. Location, on the other hand, turns out to be important; apparently at least some national raters consider location a relevant aspect of a hotel's quality.

The most important result in column 6 is that the addition of satisfaction scores does not remove all explanatory power from the country dummies.¹⁹ This is suggestive of systematic differences in national rating standards across countries. In sharp contrast, country effects are negligible in the operator rating functions. We interpret this as evidence that operators are successful in substantially mitigating the information problem by providing ratings that are more suitable for international comparisons.

How similar are the two operators' rating functions and how different are they compared to the national rating process? One way to answer that question is to compare estimated coefficients across raters. This can be done in two ways. One is to use Seemingly Unrelated Regression (SUR), which involves stacking the data and estimating the three equations jointly. A second approach is to estimate each equation separately and derive the simultaneous distribution of the estimators from the different models.²⁰ We opted for the latter method because it can more easily accommodate ordered dependent variables. We tested the null hypotheses that estimated coefficients on (i) hotel characteristics and (ii) country dummies are equal across each pair of regressions. The test statistics and *p*-values are reported in Table 7. In the first test the null could not be rejected in any of the three pairings, suggesting that all raters assess characteristics in similar ways. In the second test the null is easily rejected (at better than the 1% significance level) in the comparison between each operator and the national regression. It is rejected at the 4% significance level when the two operators are compared. Therefore the operators' rating functions are more similar to each other than to the national rating function.²¹

The similarity of the rating functions can be further explored by evaluating their predictive ability. We used each operator's estimated rating function to predict ratings for hotels that are in

¹⁹ To provide a direct comparison, we re-ran the specification reported in column 1 on the 172 observations used in column 6. The χ^2 statistic for the test $\beta_d=0$ was 41.76, only slightly higher than the 39.39 obtained in column 6. We further experimented by adding the operator rating to the national rating regression. Country dummies survived that addition also ($\chi^2=20.96$, *p*-value 0.007).

²⁰ This methodology has been proposed by Clogg, Petkova, and Haritou (1995) and can be implemented in Stata with the *suest* command.

²¹ As a check, we performed the same tests using SUR and obtained the same results.

the other operator's catalogue and then compared those predictions to actual ratings by the national rater and by the other operator. That is, for Thomson hotels we used the Thomas Cook rating function to predict the ratings Thomas Cook might have given to these hotels. The correlation between the predicted ratings and Thomson's ratings was 0.513, while the correlation between the predicted ratings and the national ratings was 0.474. Repeating this exercise for Thomas Cook hotels produced the corresponding correlations 0.461 and 0.417. In both cases predicted ratings are closer to the alternative operator's ratings than to those of the national rater. This constitutes further evidence that the ratings of the two operators are closer to each other than to national ratings.

4.4. Different standards or different assessments?

Thus far we have shown that tour operators correct country-specific idiosyncracies in the assessment of hotels and provide ratings that are comparable across countries. This does not necessarily imply that national ratings are wrong. It could be that national raters and operators are in complete agreement regarding the relative quality of each hotel but they draw the line separating the star categories at different places. In order to be concrete, consider a simple example where two raters are asked to rate four objects w , x , y , and z as being either low, medium or high quality. Each rater assess every object's quality by placing it somewhere on the quality line, as depicted in Fig. 2. He then sets the two cut-off points (denoted by double vertical lines “||”) that separate low, medium and high quality objects. In panel I of Fig. 2 the two raters make identical quality assessments; that is, they place objects w , x , y and z on the same spot on the quality line. On the other hand, they apply different standards; rater A has higher standards than rater B. As a result they both rate w as low quality, x as medium quality and z as high quality, but they disagree on y : rater A thinks y is medium quality while rater B thinks it is high quality.

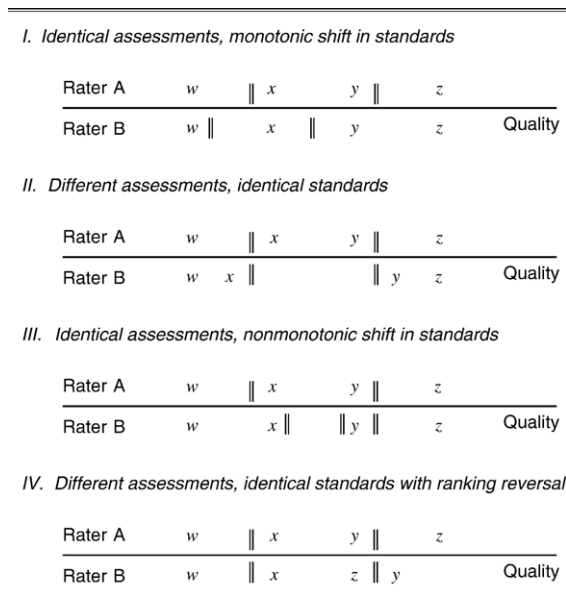


Fig. 2. A depiction of quality assessments and standards.

Table 8
Extent and direction of differences in hotel ratings

	CONFORMITY		ASYMMETRY		REVERSALS	
	TUI	TC	TUI	TC	TUI	TC
Balearics	0.632	0.564	0.920	0.765	1/3	1/2
Canaries	0.667	0.674	0.571	0.200	0/1	2/2
Cyprus	0.622	0.381	0.714	0.692	1/3	1/2
Greece	0.717	0.730	0.176	0.000	2/3	1/2
Italy	0.514	0.708	−0.059	−0.429	0/2	1/1
Malta	0.923	0.600	−1.000	−1.000	0/2	0/1
Spain	0.873	0.723	0.250	0.385	2/3	1/1
Tunisia	0.542	0.391	−0.636	−0.286	2/3	1/2
Turkey	0.867	0.727	0.000	−0.333	0/2	1/2
Mean	0.706	0.611	0.104	−0.001		
Sum					8/22	9/15

CONFORMITY=proportion of hotels that get the same rating from both raters.

ASYMMETRY=(# upgraded−# downgraded)/(# upgraded+# downgraded).

REVERSALS=number of category pairs for which ranking reversals are observed/number of pairs for which ranking reversals could be observed.

Regardless of what the “right” standard is, the existence of two raters benefits consumers. The existence of rater B helps consumers distinguish between x and y ; the existence of rater A helps them distinguish between y and z . Thus having two raters can be useful even if their quality assessments are identical, as long as they apply different standards.

The empirical implication of the above scenario is that all differences in ratings between A and B will be in the same direction. It will never be the case that rater B will give a higher rating for one object and a lower rating for another. Our data can tell us whether this is the case. If all differences in ratings are due to different standards then in each country we should observe either only upgradings or only downgradings but not both. For the purpose of this exercise we turn to our expanded dataset of 610 establishments. Rating differences for these hotels are summarized in Table 8. For each country and tour operator we constructed three summary measures that quantify the degree and direction of rating differences. CONFORMITY is simply the proportion of hotels that get the same rating from both the national rater and the operator. The average degree of conformity is 70.6% and 61.1% for Thomson (TUI) and TC respectively, with wide ranges of 51.4%–92.3% and 38.1%–73.0%. ASYMMETRY measures the direction of rating differences. It is defined as the difference between upgradings and downgradings as a proportion of the sum of the two. In countries with a positive (negative) number hotels are more likely to be upgraded (downgraded); in countries with a statistic near zero hotels are roughly equally likely to be upgraded and downgraded. As we had already seen in Table 3, Thomson is more likely than Thomas Cook to upgrade hotels. Other than that, the ASYMMETRY index is remarkably similar across operators. Hotels in Cyprus and the three Spanish regions are much more likely to be upgraded rather than downgraded, while the opposite is true of hotels in Malta and Tunisia. This suggests that national rating agencies in Cyprus and Spain are stricter in awarding stars than their counterparts in Malta and Tunisia. Only in the case of Malta, however, can differences in standards explain all rating differences. In every other country there is both upgrading and downgrading.

An obvious way for both upgradings and downgradings to arise is through differences in quality assessments, as illustrated in panel II of Fig. 2. Here rater B assesses object x lower and

object y higher than rater A.²² For instance, when the presence of a disco is reason for downgrading a hotel by rater A and upgrading it by rater B, then one can conclude that the two raters do not have a similar view of quality in relation to this hotel characteristic. Differences in assessments thus generate simultaneous upgrading and downgrading. It is also possible, however, for rating differences to arise due to differences in standards alone. In order for this to happen the differences in standards must be nonmonotonic, as depicted in panel III of Fig. 2, where rater A has a higher standard for high quality but a lower standard for medium quality than rater B. For example, rater A downgrades 3-star hotels without a satellite TV and does not upgrade them when they have a hairdresser; whereas rater B downgrades 3-star hotels only when they do not have a standard TV and upgrades them when they have a hairdresser. In this case a 3-star hotel with a standard TV and a hairdresser will be downgraded by A and upgraded by B.²³

In all situations above the two ratings are consistent with each other in the sense that there is no conflict in the implicit ordering of hotels: both raters agree that z is better than y which is better than x which is better than w . The only way to unambiguously conclude that raters assess objects differently is to observe inconsistent ratings. These may be produced in a situation like the one in panel IV, where the two raters rank the pair y and z in reverse order. Such a situation can only arise if the raters differ in their quality assessments, as differences in standards alone cannot produce a ranking reversal. Using the examples above a ranking reversal can occur when rater B considers discos a source of entertainment whereas rater A a source of nuisance. Then, a 4-star hotel with a disco can be downgraded to 3-stars by rater A whereas a 3-stars hotel with a disco can be upgraded to 4-stars by rater B. A ranking reversal cannot happen, however, when rater A considers the absence of satellite TV and rater B the absence of a standard TV as reason to downgrade a hotel; or when A does not upgrade hotels with a hairdresser whereas B does.

A ranking reversal is observed at the level of adjacent category levels. Consider for example the 3-star and 4-star categories. A ranking reversal for this category pair occurs if at least one hotel that gets three-stars from the national rater gets four-stars from the operator and at the same time at least one other hotel gets four-stars from the national rater and three from the operator. The maximum number of ranking reversals we can observe in each country is equal to the number of adjacent category pairs and varies between one and three because not all categories are observed in every country. For example, Thomas Cook does not list any 5-star hotels in the Balearics, hence there can not be a ranking reversal involving a downgrading of a 5-star hotel. In the Canaries, Thomson only lists 3- and 4-star hotels, so the maximum number of ranking reversals is one. Thus in presenting our REVERSALS measure in column three of Table 8 we provide two numbers: the first is the number of ranking reversals we observe and the second is the maximum number of reversals we could observe. The only country for which no ranking reversals are observed is Malta. This is because in Malta we only observe downgradings, so we can not have ranking reversals. Among the remaining eight regions Thomson features ranking reversals in five while Thomas Cook in all eight. Thomson has fewer reversals because it tends to upgrade more than downgrade. Collectively, the evidence suggests that the two operators do more than just

²² We illustrate the case of identical standards; clearly the same patterns can emerge if both standards and assessments differ.

²³ Although the latter scenario is a theoretical possibility, we think that the situation in panel II is empirically more plausible than the one in panel III. Nonmonotonic shifts in standards tend to make some categories wider and others narrower. The greater these shifts are, the less meaningful the ratings become. It is unlikely that this occurs on a significant scale in practice. Hence the presence of both upgrading and downgrading may be interpreted as indirect evidence of differences in quality assessment.

streamline standards across countries. They also incorporate additional information in their ratings which may lead to changes in the overall quality ranking.

In conclusion, the analysis in this section shows that differences between the national and operator rankings can arise due to a number of reasons, not all of which can be given unambiguous interpretation. In particular, upgradings and downgradings maintaining the initial hotel ordering in the two ratings can arise due to differences in standards. If this were the case, the operator ratings would be said to do no more than just streamlining standards across countries. The presence of reversals, however, suggests that the operators assess quality in a way different than the national raters; and, by incorporating additional information, their ratings can be argued to also correct the quality ranking of hotels within-countries themselves.

5. Concluding remarks

In this paper we examine the role of tour operators as information providers in the international travel industry. We employ a unique dataset which includes hotel ratings provided by both private intermediaries and by public agencies in several countries. The ratings differ about one-third of the time. We show that rating differences stem partly from the fact that each national rating agency uses its own set of standards in assigning hotels into star categories. This results in hotel ratings that are not comparable across countries. Tour operators, on the other hand, are found to apply a uniform standard to all hotels regardless of their location and thus provide ratings that are comparable across countries.

The use of different standards by tour operators and national agencies could also fully explain rating differences within a country, as long as the quality rankings of hotels are not reversed. We document instances of ranking reversals in six out of the seven countries in our sample. This suggests that, in addition to nonuniform standards, rating differences also reflect differences in quality assessment. Tour operators are in a position to accumulate specialized knowledge about hotels they rate because they are typically involved in long-term relationships with them. They also have access to hotel customers whose opinions they actively solicit. Such information is not available to national rating agencies, who are restricted to using objective and easily quantifiable criteria. Their significant informational advantage allows tour operators to provide ratings that reflect quality more accurately than ratings provided by national agencies.

We also show that prices correlate much more closely with operator ratings rather than with national ratings. The result itself is not surprising since operators set prices. It becomes more meaningful, however, when viewed in conjunction with the evidence that operator ratings incorporate more information and are thus better indicators of quality. In the absence of operator ratings, it is quite likely that prices would correlate more closely with national ratings. Therefore prices would not reflect quality as well as they do in the presence of operator ratings. The value of improved information provision by operators lies in that it provides a better matching of price with quality and therefore to a more efficient market outcome.

If UK tour operators are more effective than national agencies in providing information to prospective tourists, is there any point in governments staying in the rating business? For all the shortcomings of national rating agencies, we think that the answer is still affirmative. Tour operators' ratings are not complete – they only rate hotels they transact with – and they are generally only accessible to their UK customers (although the internet is changing that). Smaller markets may not be able to support intermediaries because the rents to be extracted are small. Improved information provision may serve well governments aiming to attract tourism from new countries that do not have well-developed tourism intermediaries.

References

- Akerlof, G.A., 1970. The market for “lemons”: quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488–500.
- Allen, F., 1984. Reputation and product quality. *RAND Journal of Economics* 15, 311–327.
- Biglaiser, G., 1993. Middlemen as experts. *RAND Journal of Economics* 24, 212–223.
- Biglaiser, G., Friedman, J., 1994. Middlemen as guarantors of quality. *International Journal of Industrial Organization* 12, 509–531.
- Chu, W., Chu, W., 1994. Signaling quality by selling through a reputable retailer: an example of renting the reputation of another agent. *Marketing Science* 13, 177–189.
- Clogg, C.C., Petkova, E., Haritou, A., 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100, 1261–1293.
- Dewally, M., Ederington, L., 2006. Reputation, certification, warranties, and information as remedies for seller-buyer information asymmetries: lessons from the online comic book market. *Journal of Business* 79, 693–729.
- European Commission, 1999. Commission Decision Case No IV/M.1524 — Airtours/First Choice. (http://europa.eu.int/comm/competition/mergers/cases/decisions/m1524_en.pdf).
- Fingleton, J., 1997. Competition among middlemen when buyers and sellers can trade directly. *Journal of Industrial Economics* 45, 405–427.
- Jin, G.Z., Kato, A., in press. Price, Quality and Reputation: Evidence From An Online Field Experiment. *RAND Journal of Economics*.
- Jin, G.Z., Leslie, P., 2003. The effect of information on product quality: evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics* 118, 409–451.
- Jin, G.Z., Leslie, P., 2006. Reputational Incentives for Restaurant Hygiene. Unpublished manuscript.
- Klein, B., Leffler, K.B., 1981. The role of market forces in assuring contractual performance. *Journal of Political Economy* 89, 615–641.
- Kliger, D., Sarig, O., 2000. The information value of bond ratings. *Journal of Finance* 55, 2879–2902.
- Marvel, H.P., McCafferty, S., 1984. Resale price maintenance and quality certification. *RAND Journal of Economics* 15, 346–359.
- Monopolies and Mergers Commission, 1997. Foreign Package Holidays: A report on the supply in the UK of tour operators’ services and travel agents’ services in relation to foreign package holidays. http://www.competition-commission.org.uk/rep_pub/reports/1997/412foreign.htm.
- Rubinstein, A., Wolinsky, A., 1987. Middlemen. *Quarterly Journal of Economics* 102, 581–594.
- Shapiro, C., 1983. Premiums for high quality products as returns to reputations. *Quarterly Journal of Economics* 98, 659–680.