Nicolas Gothelf (Belgium), Denis Herbaux (Belgium), Vincenzo Verardi (Belgium)

Do theme parks deserve their success?

Abstract

This paper, starts by building a typology of theme parks, based on the satisfaction of consumers and the degree of "theming" using a dataset of 70 European parks. Then, controlling for the presence of outliers, the authors a ranking of parks with respect to prices and attendance. The article finally points out the main characteristics that impact theme parks attendance and entry prices. They stress the positive correlation between prices and attendance and highlight the important underlying role of investments in this relation.

Keywords: clustering, theme parks, outlier identification.

Introduction

Theme parks have been a part of the American dream for the past decades. Among the competitors, the natural superstar, the Disney Company, brings the idea of magic and fairy tales since 1957 for the happiness of young and old. Nowadays, the supply has diversified to a wide class of themes (e.g., aquatic, movies, stunts, etc.) as well as the type of public who is aimed at. Of course, all this magic and broadly all, leisure facilities, has a non-negligible cost to the household. One relevant question that arises is the relation between these costs and what is offered by a theme park. Another question is the relation between the level of attendance and the effective offer of the park. In other words, does the theme park deserve its reputation based on concepts that we develop further such as its *quality* and *theming*.

An immediate issue that arises is the definition of a theme park in order to set the perimeter of the subsequent analysis. A theme park or amusement park is a generic term for a collection of rides and attractions assembled for the purpose of entertaining a large group of people. Historically, periodic fairs, such as the Bartholomew Fair which began in England in 1133, are the ancestors for the modern amusement park. Beginning in the Elizabethan period, the fair had evolved into a center of amusement with entertainment, food, games, and carnival-like freak-show attractions (Judith Adams, 1991). The seasonal celebration was a natural place for development of amusement attractions. In Europe for instance, the Oktoberfest is not only a beer festival but also provided amusement park features as far back as 1810, when the first event was held in Munich, Germany. In the United States, county and state fairs also played a part in the history of amusement parks (Samuelson and Yegoiants, 2001). These were annual events that were usually held for a short period, a week or two, to celebrate a good harvest. These fairs featured livestock exhibits, baking and cooking competitions (Judy Alter, 1997).

Modern theme parks turn out to be more elaborate than a simple city park or playground and are usually provide attractions meant to cater adults, teenagers, and small children. Furthermore, a theme park is a type of amusement park which was built around one or more themes, such as an American West theme, or a jungle theme.

An additional but crucial element to add to the definition is the "pay-one-price" characteristic in comparison with "pay-as-you-go" funfairs that are excluded from our analysis. The "pay-one-price" ticket was first introduced by George Tilyou at the Steeplechase Park, Coney Island in 1897 (Judith Adams, 1991). Note that at the beginning this feature was not immediately present even in today's well-known parks. For instance, Disneyland opened using a "pay-as-you-go" format and guests paid ride admission fees at each attraction. Angus Wynne, founder of Six Flags Over Texas, first visited Disneyland in 1959 and noted the park's "pay-as-you-go" format should turn to "pay-one-price", as he did for his park. On the one hand, the main disadvantage of "pay-one-price" is that guests may not attend a park if they don't believe they'll get their money's worth but on the other hand, "pay-one-price" parks imply that guests can more easily budget their visit and may be more likely to experience an attraction they've already paid for. In our analysis, we focus on parks offering this "pay-one-price" feature.

The concept of a theme park have been clarified, we turn to a brief description of the recent evolution and present situation. Approximately 225 parks existed worldwide in 1990. The total number of visitors was around 300 million (for a global turnover of 7 billion USD). The sector increased to 340 parks with 545 million visitors and a turnover of 13.8 billion USD in 1999 (Weiermair and Mathies, 2004). Though already big, the park industry is still growing. Indeed, on average, the attendance has grown by approximately 4% for the period of 2005-2007 for top attendance parks.

As reported in the 2007 Themed Entertainment Association/Economics Research Associates (TEA/ERA, 2006; 2007; 2008), main playgrounds for entertainment parks are North America (almost 45% of the global attendance), Europe and Asia (almost one quarter each). Note, however, that international entertainment firms are currently developing huge projects of theme parks in the Middle East. Top attendance theme parks in these three regions (North America, Europe and Asia) sum up to 280 million visitors. This last figure corresponds to the attendance of the "big ten" theme park attraction companies: Walt Disney Attractions, Merlin Entertainment Group, Universal Studios Recreation Group, Six Flags Inc., Busch Entertainment, Cedar Fair Entertainment Company, Parques Reunidos, Compagnie Des Alpes (CDA), Herschend Family Entertainment, Everland.

Several studies have been carried out to analyze the different dimensions of experience from a theme park and the various benefits, as well as factors having influenced them, for instance (Bigné J. et al., 2005). Fewer have discussed the economic aspects such as in (Anton-Clavé, 2004; Wanhill, 2008). An often-cited reference is (McClung, 1991) that focuses on factors influencing attendance to theme parks consumer choice in the US. However, the American market differs in several dimensions from the European (and the Asian) market. While the population is approximately the same in North America and in Western Europe, overall attendance is quite different. In North America, the top twenty parks in terms of attendance sums up to 120 million, while the European top twenty reaches only 60 million visitors. On top of this difference of attendance, there also exists dissimilarity in terms of localization. In North America, the top thirteen parks in terms of attendance (more or less 100 million visitors per year) are located in two states, namely California and Florida. In Western Europe, the top thirteen parks are spread over eight different countries.

A third difference between Europe and North America is the structure of the market. Five major companies are operating in the US: Walt Disney, Universal, Busch, Six Flags and Cedar (Swarbrooke, 2002). These companies can be divided into three types of parks. First, there are original and highly themed parks, owned by three companies, namely Walt Disney, Universal and Busch that are mainly located in California and Florida. The concentration of parks in California and Florida seems to indicate that part of the American market can rely on visitors that are ready to travel long distances and stay in parks for long periods. This behavior is not really observed in Europe, except maybe for Disneyland Paris, the first touristic destination in Europe. Second, Six Flags Inc. develops a strategy that consists in creating (or updating) homogeneous theme parks outside big cities all over the US. These parks are mainly composed of roller coasters, some theming (e.g., a Looney Tunes theme for a children's area in the parks), and sometimes a waterpark. Finally, there also exist many local parks the attendance of which is small in comparison with the "big five". The European market is mainly characterized by parks created by small and local investors and then sometimes acquired by big companies that do not change their initial characteristics. One could argue that causes for the differences observed between the two regions are, besides cultural differences, the existence of multiple languages in Europe as well as high moving costs. However, the increasing number of low cost airlines, the existence of the euro as the single currency in many European countries as well as the increasing number of theme park websites in multiple languages seems to make these arguments less relevant. The typology of parks in Europe below helps us to point out this diversity and to grasp the idiosyncrasies of the European logic. The main questions we raise are what are the determinants of entry prices and attendance and which characteristics can explain these differences?

The paper is organized as follows. Section 1 describes dataset and variables. A typology is proposed in Section 2. The model and the hedonic pricing are examined in Section 3. The last Section concludes.

1. Dataset and variables

Our dataset includes 70 theme parks, located in 12 European countries. These parks opened between 1843 (Tivoli Gardens, Denmark) and 2003 (Movieland Studio, Italy). These 70 parks have been selected in the dataset on the basis of attendance. Indeed, only parks with an attendance larger than 100,000 visitors a year have been considered. Attendance of parks in our database ranges from 100,000 visitors per year (Happyland New, Switzerland) to 12,000,000 (Disneyland Paris, France). For each park, we collected a set of characteristics that are presented hereunder. For the sake of clarity we present them as bulleted items. Variables that are often used in the related literature are: short waiting time, good climate/environment, proximity or family atmosphere (Moutinho, 1988).

1.1. General information about variables. There are several variables:

- Attperweek: this variable gives the annual attendance of the theme park. It mainly comes from the TEA/ERA Theme Park Attendance Report for 2007 and from companies' annual report. When data were not available in the report, we completed it using local press articles. To standardize for opening periods, we normalize attendance by the number of weeks per year the park is open. Taking opening weeks into account, the standardized variable ranges from 4,167 to 230,769, with median 31,660.

- Total: this variable summarizes the total number of attractions and the number of shows available in each park. This gives an idea of the size of the theme park. The smallest park counts 11 attractions and shows (Sommerland Syd, Denmark), while the largest has 59 attractions and shows (Disneyland Paris, France).
- PA: is the entry price for one day for an adult. Minimum price is 11 euros (Wild und Freitzeitpark Klotten, Germany) and maximum price is 49 euros (Disneyland Paris, France). Mean price is 27 euros.
- Age: is a normalized value of how old a park is. 0 is the value for the oldest park and 1 for the most recent. It is calculated by taking the year the park was created and substracting the year the first park was created (i.e., 1843). This new variable is then divided by the year the last park was created (i.e., 2003).
- Population: this variable measures how many people live in the area surrounding the park within a 150 km radius that corresponds door to door to a travel of 2 hours at most. The idea, developed in Section 3, is that everything else being equal, a park surrounded by more people should have a higher attendance.

Except for attperweek and population, all data have been collected on individual theme park websites.

1.2. Specific characteristic variables. Specific characteristics of parks can also be identified. Indeed each park has a specific vocation and its attractions are not all of the same type. To remain brief, we can say that four groups of type of attractions can be identified:

1. Thrills: thrill rides are mainly designed for teenagers and young adults who are looking for adrenaline. Examples of thrill rides are roller coasters, top spin or free-fall towers.
2. Family: family rides are conceived to allow the whole family to enjoy these attractions. Classical examples are Dark Rides and Big Wheels.
3. Kiddies: are attractions exclusively restricted to small children.
4. Waters: this category is more heterogeneous than the previous ones in the sense that water rides can often also be classified as thrills (flume rides or splash) or family (tow boat rides). For simplicity, as soon as an attraction involves water, except for dark rides, we counted it in the waters category.

Depending on the question investigated, we will use the classification of attractions in absolute terms (as described here above) or in proportion of the total number of attractions:

- Propthrills: is the proportion of thrill attractions on the total number of attractions.
- Propkiddies: is the proportion of kiddie attractions on the total number of attractions.
- Propfamily: is the proportion of family (that is non-thrill and non-kiddie) attractions on the total number of attractions.

These three variables allow us to directly characterize the park in terms of targeted market by observing which kinds of attractions are mostly offered by the park.

The last two characteristics should be considered to be related to the perceived satisfaction from the consumers and the investments made by the companies. These are:

- Quality: measures the quality of the park and ranges from 0 to 10. It has been constructed using seven theme park reviews available on the web. This includes four general review (including attractions, food, services, etc.), two coaster reviews (2007 Steel and Wooden Roller Coaster Poll, and Coaster Force review), and one review of the top 10 best attractions (including top 10 best rides, best 3D Show, Best Bobsled Coaster, Best Darkride, Best Freefall, Best Haunted House, Best Indoor Coaster, Best Interactive Darkride, Best Inverted Coaster, Best Log Flume, Best Madhouse, Best Minetrain Rollercoaster, Best Shot'n Drop, Best Show, Best Simulator, Best Spinning Wild Mouse, Best Steel Roller Coaster, Best Stunt Show, Best Water Coaster, Best Watersplash, Best Wild Mouse Coaster, Best Wild-water ride, Best Wooden Coaster). Quality is calculated as the average of the scores obtained by each park in the seven reviews. These reviews are well-known in the park fan community. Even if those who fill up these reviews are not representative of the whole population, they can be seen as experts in the field of theme parks. Moreover, among them are teenagers, grownups, singles, males, females, parents, etc. Hence, the variable quality is a subjective variable of perceived quality from experts. Quality will allow us to check to what extent the attendance of a park follows the experts' point of view. Ordering theme parks by this sole criterion leads the following results. The best parks in terms of quality are, in order, Phantasialiand (Germany), Efteling (The Netherlands), Disneyland Paris (France), Europa Park (Germany) and Walibi World (The Netherlands).
- Theme: this variable summarizes the degree of theming of the park ranged from 0 to 10. It is calculated as a weighted sum of the proportion of themed attraction in the park (a themed attraction is an attraction with a particular story and/or a particular set), the fact that the park has a well defined theme or image (the park is represented by a specific character or the park as a well recognizable theme), and the proportion of themed areas (areas designed with a specific set or having a specific

character as theme). For example, Alton Towers (UK) has no general image, but is divided into well themed areas, while the opposite is true for Legoland Billund (Germany). All attractions are themed in Disneyland Paris, while none are in Happiland New. Conversely with the previous variable, is purely objective. The five best parks in terms of theme are: Disneyland Paris, Walt Disney Studio (France), Phantasialand (Germany), Movie Park Germany (Germany) and Parque Warner (Spain).

Table 1. Summary statistics

	Mean	St. dev.	Pp 50	Min	Max
Age	0.80	0.19	0.82	0.00	1.00
Attperweek	41934.19	39893.07	31660.50	4167.00	2.3e+05
Total	33.24	9.98	32.00	11.00	59.00
PA	26.96	8.16	27.00	11.00	49.00
Population	8.4e+06	5.8e+06	7.0e+06	1.3e+06	2.9e+07
Propkiddies	0.31	0.10	0.32	0.00	0.51
Propthrills	0.19	0.07	0.19	0.07	0.33
Propwaters	0.10	0.05	0.10	0.00	0.25
Propfamily	0.40	0.10	0.39	0.22	0.76
Quality	4.94	1.54	4.48	3.20	9.43
Theme	3.12	3.09	1.48	0.00	10.00

In the next Section, we normalize quality, theme and total from 0 to 1 scale (that we call n-quality, n-theme and n-total[1]. Attperweek, corrected for opening weeks, and population are taken in logarithm.

2. Typology of parks

The first step of our analysis is the creation of a typology of European parks. In the literature, Fodness and Milner (1992) proposes a cluster analysis focusing on theme park consumers characteristics. Conversely, the goal of this Section is to understand the logic behind Europe's theme parks based on the parks' characteristics and the satisfaction of consumers. To do so, we use some standard clustering techniques on the following variables: theme, quality, total (all three being normalized), and propkiddies, propthrills, and propfamily.

For the cluster analysis, we start with a preliminary complete linkage dendrogram based on a L2 dissimilarity measure. The dendrogram, that is available in Appendix A, clearly shows that four main clusters emerge. We, therefore, continue our clustering analysis by calling on a k-means algorithm with four centers. To allow for the k-means code to remain stable, we ask the program to start with the following four centers: Europa Park, Bellewaerde, Paulton Leisure Park and Legoland Deutschland. As can be seen in the dendrogram, these parks are clearly located in the center of the four main clusters identified in the preliminary analysis.

We, thus, end up with 4 stable groups, indexed from 1 to 4. These groups are represented in Figure 1, which has to be read in the following way. On the horizontal axis, one finds the level of theming, and on the vertical axis, the level of quality. The centered number indicates the cluster. The size of the hollow circle of each park represents the size of the park in terms of the number of rides. Finally, the diamond-hollowed number indicates the targeted market to be more kiddie than family oriented.

Five facts can be extracted from these results:

- A first finding is that there seems to be no specific correlation between quality and theme in European parks. Disneyland Paris, Efteling, Europa Park, Parque Warner, Phantasialand are of high quality and highly themed. One group of parks composed by Fraispertuis City, Movie Park Germany, Parc Asterix and Walt Disney Studio is highly themed but of average quality. Besides, the group formed by Plopsaland and Terra Mitica, and Movieland Studio contains parks that are relatively well themed but not recognized as good quality parks. Furthermore, many parks are almost not themed at all, but vary in terms of quality. Two groups (one composed by Bobbejaanland, Gardaland, Liseberg, Thorpe Park and Walibi World; and the other by Heide Park and Pleasure Beach) are of high quality and low theme. There are also a lot of small parks of low quality and low theme.
- Second, many parks and groups of parks are kiddie oriented. Among them, a lot exhibit low theme and low quality. The main reason for this is that kiddie rides are mostly simple fair attractions. Moreover, these parks are also relatively small. This can be explained by the fact that either owners of such parks design them explicitly for small children, without any other objective, or these parks are planned to grow in the future, and that first investments made are the cheapest possible, which are typically the case for kiddie rides.

[1] The formula used is $\frac{x_i - \min\{x\}}{\text{Max}\{x\} - \min\{x\}}$.

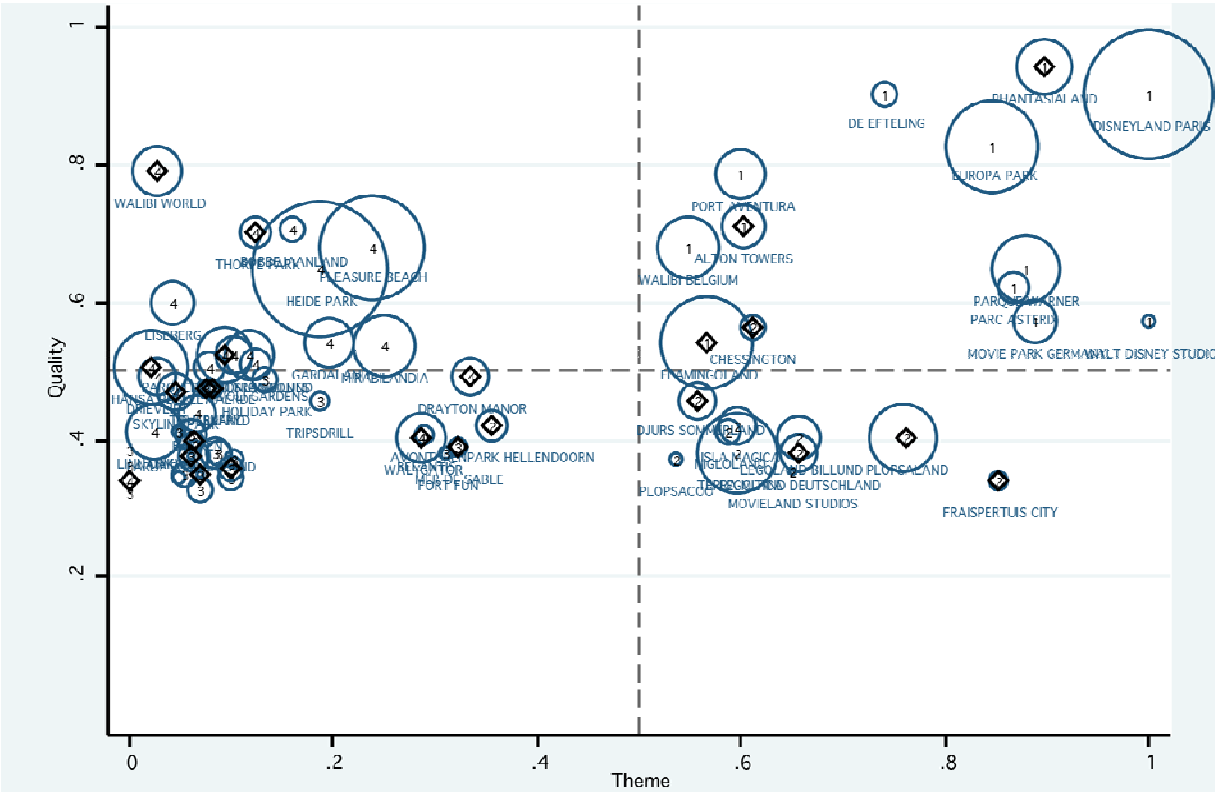

Fig. 1. Typology of parks

- Third, no park is specifically thrill oriented conversely to the US (Six Flags parks for instance). European parks always embodied a larger kiddie or family component. Parks from cluster 1 and 4, which are of higher quality, exhibit a larger proportion of thrill rides. This could be partially explained by the bias one can expect from sources on the basis of which we built the quality variable[1].
- Fourth, two parks exhibit high quality and high theme, and are family oriented (Efteling and Disneyland Paris). This result is perfectly in line with the general opinion inside the park fan community, since both parks are recognized as being among the "best". Moreover, the similarities of these two parks are not really surprising given that Walt Disney decided to create the first Disneyland (in Anaheim, California) after visiting Efteling, meaning that the concept and characteristics of both parks are similar.
- Fifth, one group is of particular interest. This group is composed by Europa Park, Parque Warner, Phantasialand and Port Aventura. This group does not seem to be oriented toward a specific type of market. Indeed, the values of the centered and normalized prop-variables are almost equal to zero. This can be interpreted as the fact that these parks are designed for all profiles of consumers.

We can then conclude that these parks are "general parks". Moreover, these parks all exhibit relatively similar patterns of theming: a general theme more or less defined, but lands or subsections of the park and attractions with a high level of theming.

We also compute which type of parks is the most present in each country. Results are the following: Switzerland, Germany and the Netherlands have a majority of family oriented parks. This is coherent since these countries are relatively close from a cultural point of view.

Denmark and Finland are both more kiddie oriented, the same is true for France. Finally, Belgium, Spain, Italy, Norway, Sweden, and the United Kingdom are more thrill oriented. For the last three countries, this can be explained by some cultural proximity with the United States, which are also known as thrill oriented. The similarity between Spain and Italy is not surprising either.

3. Model

As stated in the introduction, the main objective of the paper is to determine which the characteristics of entry price and attendance are. One relevant question is of course to find out if price and attendance are related. Figure 2 presents a simple scatter plot of the entry price against the attendance per week (both in log) for each park.

As can be seen, both variables are strongly and positively correlated. It is quite puzzling to observe that parks with the highest attendance also exhibit higher price. This could mean that high prices attract people. The answer to this puzzle is in fact simple. We believe that entry prices depend on the characteristics of parks (and hence on the investment made for

[1] We control this in the regression Section.

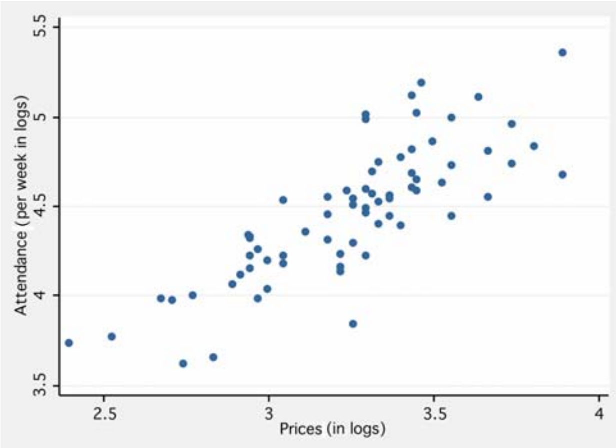

Fig. 2. Attendance vs. price

the park), and not really on attendance. Some anecdotic evidence supports this point. For example, interviewed in 2006, one of the directors of CDA parks stated that "There is no reason to increase the price by more than inflation in the absence of large investments". This means that, contrary to most traditional goods of which the price decreases with time, a theme park's entry price is more or less stable when corrected for inflation, and "jumps" only occur when investments are made. To illustrate this point, it is interesting to look at some additional evidence. For example, in 2006, Phantasialand increased its price by 16% when it opened the "Black Mamba", an inverted roller coaster which is considered by the park fan community as one of the best roller coasters in Europe. That year, attendance increased by 19%. In 2007, Port Aventura raised its price by 8% and opened "Furius Baco", a huge launch coaster with a totally new train design, which involved a 6% increase in attendance. While its price stayed fixed between 2002 and 2005 and no significant investment was made, Disneyland Paris engaged in a 250 million euros investment plan for the period of 2005-2009. During that period, price increased on average by 5% yearly (plus inflation) and attendance increased by 8% on yearly average. Plopsaland (Belgium) and Europa park (Germany), after a long period without increasing their price significantly, both raised their price by 7% when opening a new coaster (note that price variations are again considered while controlling for inflation). A final point shows that prices are not really affected by attendance (at least in the shortrun) is that the price is set before the beginning of the season and it is almost never modified during the year while attendance varies. Indeed, it is really hard for firms to decrease entry price in order to attract consumers. Rebates are not frequent, and, in general, concern a limited number of tickets for a specific period of time. On top of that, rebates are often part of a package, so that it is actually not really a decrease in entry price. Remember that we use the full price here.

Having explained the latter, the positive correlation between prices and attendance is no longer a puzzle. Consumers choose parks depending on attractions, and they are willing to pay a higher price to enjoy more attractions.

To identify the determinants of price and attendance, we run two separate hedonic robust regressions on price and attendance trying to understand which characteristics affect these variables. An analysis of the residuals is then performed to identify overpriced and underpriced parks. Note that though it is not evident that prices affect attendance directly (and vice-versa), it is dangerous to state that these variables are independent from one another. To cope with this, in the final estimation we use the Seemingly Unrelated Regression estimators to fit the regression parameters (obviously after down weighting the influence of outliers). In this way we will control the fact that these two characteristics are shadowy related via the right-hand side variables in the regression that are the characteristics of parks.

3.1. Hedonic pricing. Hedonic demand is an economic theory which asserts that the price an individual is willing to pay for a good reflects the sum of the values he awards to each of the characteristics of that good. This theory hence assumes implicitly that the object of demand is the sum of individual characteristics rather than the good itself. To estimate these individual values, a common approach is to run a hedonic regression using the price of the good (generally in log) as a dependent variable and its characteristics as independent variables (assuming that these characteristics are well identified). An analysis of the residuals can then be informative in understanding if a good is sold above or below its "fair" price and, even more importantly, if individuals are outlying. Then, by recognizing the type of outlier, a more thorough analysis can be performed on individuals. Broadly speaking, three types of outliers can be identified in cross-sectional regression analysis. Rousseeuw and Leroy (1987) define them as *vertical outliers*, *bad leverage points* and *good leverage points.* To illustrate this terminology, we consider a simple linear regression as shown in Figure 3 (the generalization to higher dimensions is straightforward). For the sake of clarity, we shade the area where 95% of the observations are expected to appear in case of Gaussian data. *Vertical outliers* are those observations that are outlying in the dependent variable (the y-dimension) but are not outlying in the design space (the x-dimension). These individuals have characteristics that are comparable to the other individuals but their price is not (too high for positive residuals and too low for negative ones). *Good leverage points*, are observations that are outlying in the space of explanatory variables (i.e., have characteristics that are different from the bulk of data) but that are located close to the regression line. These observations are characterized by attributes that are very different to the other individuals but whose price is in accordance with their attributes. Finally, *bad leverage points*, are observations that are outlying in the space of explanatory variables and that are located far from the regression line. These points have characteristics that are very different from the others and prices that are either excessive (positive residuals) or insufficient (i.e., negative residuals) given their characteristics.

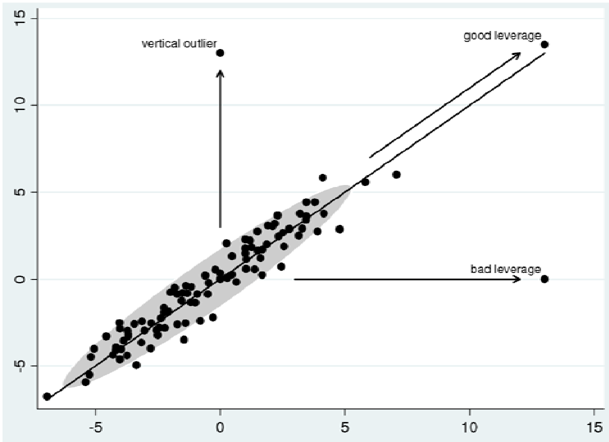

Fig. 3. Vertical outliers, good and bad leverage points

While identifying and recognizing the type of outliers is easy in a simple regression analysis (as represented in Figure 3), it is much less the case in higher dimensions. However, a simple graphical tool, proposed by Rousseeuw and Van Zomeren (1991), can be used in a multiple regression setup. This graphical tool is constructed by plotting the robust standardized residuals of the hedonic prices regression on the vertical axis which gives an idea of outlyingness in the dependent variable and plotting on the horizontal axis a measure of the (multivariate) outlyingness in the space of explanatory variables. It is important to highlight that the standardized residuals must be fitted using a robust estimation procedure since LS residuals are uninformative in case of the existence of outliers given the distortion the latter induce in the estimation of the parameters. Similarly, the degree of outlyingness in the space of explanatory variables must be based on a robust distances estimator. Finally, to visually assess which are the outlying individuals we draw a vertical line and two horizontal lines to delimit the critical values of respectively the standardized residuals (which are normally distributed) for the y-axis, and of the robust distances (which are distributed as a Chi-squared for the x-axis).

3.2. Robust regression and identification of outliers. *3.2.1. Outlyingness in the dependent variables.* Assume we want to estimate a regression model of the type:

$$y_i = X_i\theta + \varepsilon_i \text{ for } i = 1, \ldots, n, \tag{1}$$

where n is the sample size, X is the matrix of the explanatory variables, y is the dependent variable, θ is the vector of regression parameters and ε_i is the error term. Vector θ is generally estimated least squares (LS) by minimizing the sum of the squared residuals. More precisely

$$\hat{\theta}_{LS} = \arg\min_{\hat{\theta}} \sum_{i=1}^{n} r_i^2 \ , \quad \text{where } r_i = y_i - X_i\hat{\theta}, \tag{2}$$

Given that the square function awards excessive weight to large residuals, LS is extremely sensitive to extreme values and might lead to poor estimations in case of presence of outliers in the dataset. To cope with this, several alternative methods have been proposed in the statistical literature. A well-known class of estimators is that of S-estimators that combine strong robustness and good asymptotic properties.

The intuition behind this class of estimators is simple. Recall that for LS, the objective is to minimize the variance of the residuals. However, since the variance is not a robust estimator of spread, LS breaks down in the presence of outliers. The idea behind S-estimators is to minimize another measure of the dispersion of the residuals that is less sensitive to extreme values (we could for example think of the interquartile range). More precisely, S-estimators of regression are defined by:

$$\hat{\theta}_S = \arg\min_{\hat{\theta}} s(r_1(\hat{\theta}), \ldots, r_n(\hat{\theta})), \tag{3}$$

where s is a robust measure of dispersion. The robust spread is generally estimated by an M-estimator of scale which can be seen as a robustified version of the variance. The complexity associated to an M-estimator of scale clearly goes outside the scope of this paper. We prefer not to discuss it furthermore here and refer the interested reader to more technical articles (for example see Verardi, and Croux, 2009).

3.2.2. Outlyingness in the space of explanatory variables. To identify outliers in the space of the explanatory variables, a common procedure is to use Mahalanobis distance. This distance is defined as:

$d_i = \sqrt{(x_i - \mu)\Sigma^{-1}(x_i - \mu)^t}$, where μ and Σ are respectively the multivariate location vector and covariance matrix of the explanatory variables. Note that dummy variables cannot create outlyingness in the space of the explanatory variables and are therefore not included in the estimation of Mahalanobis distances. Obviously both μ and Σ should be robustly estimated if we want these distances to resist to the presence of outliers. Several methods have been proposed for this. In this paper, we use the minimum covariance determinant estimator of location and scatter, introduced by Rousseeuw (1984), that has been proven to be particularly well suited in this context. The resulting d_i converges asymptotically to a χ_p^2 (see Hardin and Rocke, 2005; Zuo et al., 2004). The cutoff generally used to identify a leverage point is hence the 95% quantile of the χ_p^2 (i.e., when the measure of outlyingness is above $\sqrt{\chi_{p;0.95}^2}$, an observation is considered as a leverage point). To provide the intuition behind the minimum covariance determinant (MCD) estimator, it is helpful to recall the notion of generalized variance. This measure, originally introduced by Wilks (1932), is a one-dimensional assessment of multidimensional scatter. For the sake of clarity, we explain this concept by calling on a 2 X 2 covariance matrix. The generalization to higher dimensions is straightforward.

Let us define a covariance matrix $\Sigma = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 \end{pmatrix}$, where $\sigma_{x_1}^2, \sigma_{x_2}^2$ and $\sigma_{x_1x_2}$ are respectively the variance of variable x_1, the variance of variable x_2 and the covariance between the two. The generalized variance is defined as the determinant of Σ (i.e., $\sigma_{x_1}^2\ \sigma_{x_2}^2 - \sigma_{x_1\ x_2}^2$). To grasp why this measure can be seen as a generalization of the variance, it is helpful to look more closely at the expression of the determinant. This expression is composed of two elements: the product of $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ and (minus) the squared covariance $\sigma_{x_1x_2}^2$. The first term ($\sigma_{x_1}^2\sigma_{x_2}^2$) represents the raw bi-dimensional spread of the observations. However, if x_1 and x_2 are not independent, each of the variables conveys some information about the other one. This redundant information should be accounted. This is done (in the case of a 2 X 2 covariance matrix) by removing the second terms (i.e., square of the covariance $\sigma_{x_1x_2}^2$). What remains is the bi-dimensional spread once the co-variation has been accounted. A similar reasoning can be adopted for higher dimensions. However, the complexity of the expression for the determinant makes it less intuitive.

Having defined the generalized variance, it is now easy to present the underlying principle of MCD. Assume we want an estimator of the covariance matrix that withstands a contamination of up to 50% of sample points[1] (it is then said to have a breakdown point of 50%). The basic idea of MCD is to identify the subsample containing 50% of the observations that is associated to the smallest generalized variance. This is equivalent to finding the subsample with the smallest multivariate spread (and thus of the more multidimensionally similar observations). The MCD robust covariance matrix, labeled Σ_{MCD} is then defined as the classical covariance matrix estimated on this subsample.

To identify such a subsample, the idea is to consider all possible subsamples containing 50% of the observations and flag the one with the smallest covariance matrix determinant. This is of course a cumbersome task. Imagine for example that we have a dataset of 70 observations.

If we want to consider all the different subsamples containing 35 observations, the total number of subsamples to check is $\binom{70}{35} = 1.12 \times 10^{20}$. Fortunately, the number of subsamples to check can be dramatically reduced by using some very efficient algorithms but this is out of the scope of this paper (for further details see Verardi and Dehon, 2010).

3.3. Results. For each equation, we use the following explanatory variables: age of the park, population in the neighborhood (as a measure of the potential attendance), the number of languages on the internet site (as a measure of international openness), the number of each kind of attractions, the number of shows, the quality and the theme variables, as well as country dummies to take into account any potential country fixed effect or cluster fixed effect (such as purchase power or cultural specificities). We also control for the grouping identified in the clustering analysis section of the paper. Figure 4 presents the results.

The left part shows outliers in terms of price, and the right part outliers in terms of attendance. Some interesting results emerge. We present in plain dots individuals that are outlying in any (or both) regression. First, there are only very mild outliers in attendance. This means that when controlling for the various characteristics of parks, each park has an attendance that fits its characteristics adequately. This is not true for the price. Indeed, when looking at the left side, one can observe that some parks exhibit a higher or a lower price given their characteristics. For example, Skyline Park, Bonbonland and Wild und Freizietpark Klotten practice a price which is lower than what could be expected. However these outliers are mild. One can also see that big European parks such as Disneyland and Walt Disney Studio, which can be said to be different from the average in terms of characteristics (given

[1] A similar reasoning can be adopted for any h% contamination where h < 50%.

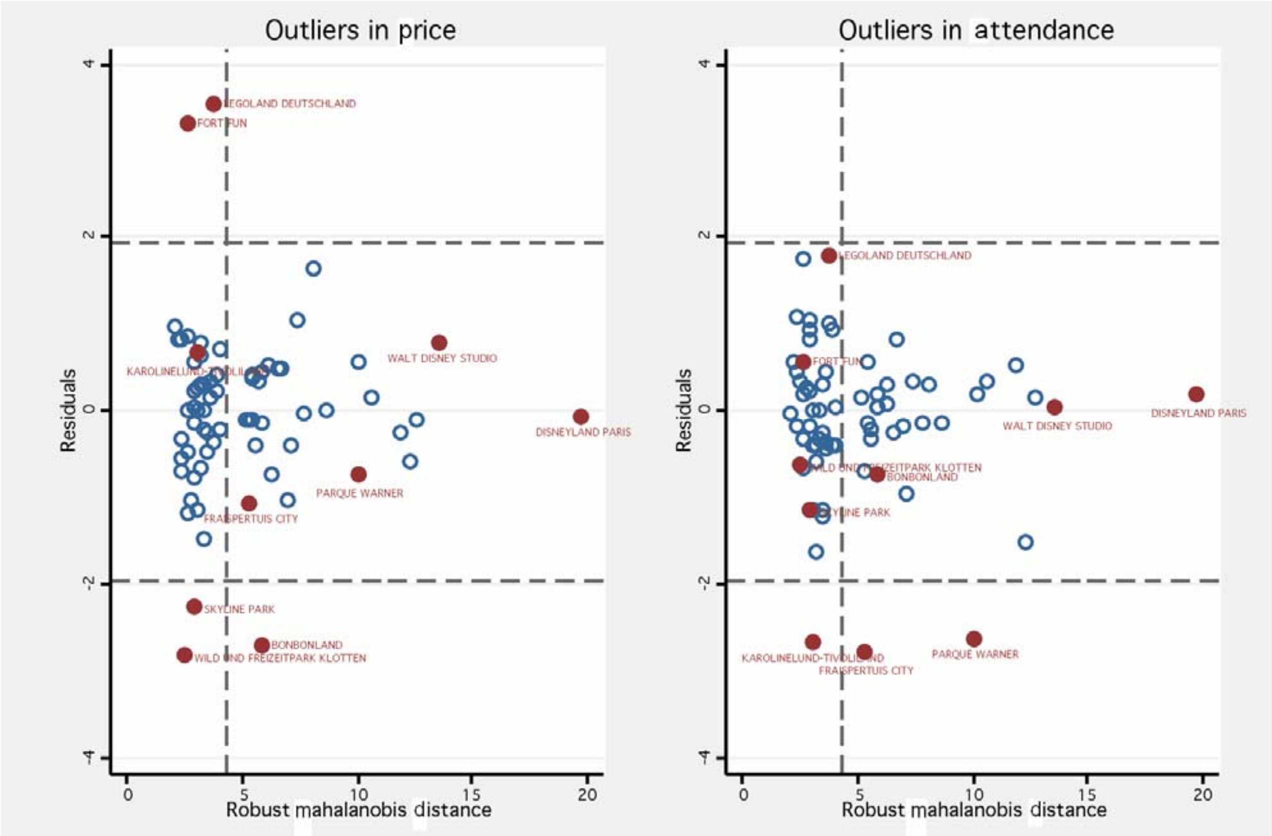

Fig. 4. Outliers in price (left) and in attendance (right)

their value of the robust Mahalanobis distance), but are neither outliers in terms of attendance, nor in terms of price. This observation (of non-outlyingness) is actually true for most parks. This in turn validates our assumption that agents choose a park depending on its characteristics, and are ready to pay the price for it, price which is fixed depending on the investments made. One particular and interesting case (and exception) appears when considering Legoland Germany and Fort Fun. Indeed, both parks can be considered as average parks, with the appropriate attendance. However, both are slightly too expensive with regard to their characteristics.

After the detection of outliers, we can now turn to the estimation step. As stated previously, the two hedonic equations are estimated as a system. The identified outliers are awarded a lower weight (which is proportional to their degree of outlyingness) in the price equation to avoid biases. In Table 2 the results for the price regression are presented.

Table 2. Price regression

Logpa	Coef.	Std. err.	P > t
Age	-0.092	0.078	0.238
Logpopulation	0.077	0.08	0.002
Languages	0.039	0.012	0.001
Kiddies	-0.002	0.003	0.664
Family	-0.007	0.004	0.023
Thrills	0.019	0.008	0.017
Waters	0.034	0.009	0
Shows	0.020	0.005	0
Quality	0.049	0.015	0.001
Theme	-0.09	0.017	0.570
Ch	-0.214	0.109	0.051
De	-0.126	0.059	0.033
Dk	0.414	0.079	0
Sp	0.184	0.067	0.006
Fi	0.486	0.101	0
Fr	0.109	0.062	0.076
It	0.011	0.080	0.885
Nl	-0.190	0.063	0.003
No	0.679	0.123	0
Uk	0.165	0.067	0.014
Se	0.065	0.101	0.515
Cluster - 2	-0.221	0.060	0
Cluster - 3	-0.251	0.114	0.028
Cluster - 4	-0.129	0.100	0.195
Cons	1.318	0.552	0.017
Number of observations	70		
R-sq	0.9057		
Chi2	595.78		

Concerning the price, amongst the various kinds of attractions one can find in a theme park, only thrill rides, waters, and shows have a significant and positive impact on the price. This suggests that these kinds of entertainments are probably more expensive to build. The level of quality also plays a major role in the formation of price which was expected. Indeed, since the quality variable was built on subjective assessment by consumers, one can expect that individuals will judge a park as being of higher quality if the rides and services are of better quality, which obviously means more expensive. Concerning country dummies, Germany and the Netherlands are cheaper (all other things being equal) than Belgium (the reference country), while Spain, Denmark, Finland and Norway are more expensive. This can simply be explained by the difference in terms of general prices existing between these countries.

One surprising result is the fact that theming does not influence prices. This can be explained in the following way: the theme variable was constructed in an objective way, simply considering whether there is some theming in the park, but without measuring its quality. This aspect is in fact taken into account in the quality variable. Hence, putting theme in a park does not influence prices except if its quality is high. Finally, it seems that a park with a higher surrounding population exhibits a higher price. This can probably be explained by a well-known feature in economic geography that is: prices are generally higher in large cities than in smaller ones. Another striking result is the negative impact of family rides on prices. The dependant variable (full-entry price for an adult) is such that an increase in family rides will not attract more families[1].

Table 3. Attendance regression

Logattperweek	Coef.	Std. err.	P>t
Age	-0.350	0.112	0.002
Logpopulation	0.225	0.110	0.041
Languages	0.011	0.017	0.519
Kiddies	-0.003	0.005	0.549
Family	0.013	0.005	0.003
Thrills	0.020	0.011	0.072
Waters	0.013	0.013	0.330
Shows	0.003	0.007	0.653
Quality	0.084	0.021	0
Theme	0.033	0.024	0.180
CH	-0.182	0.157	0.246
DE	0.056	0.085	0.507
DK	0.554	0.114	0
SP	0.361	0.096	0
FI	0.48	0.145	0.001
FR	0.202	0.088	0.022
IT	0.418	0.115	0
NL	-0.105	0.091	0.248
NO	0.630	0.177	0
SE	0.361	0.144	0.012
UK	0.148	0.097	0.127
Cluster 2	-0.053	0.086	0.543
Cluster 3	-0.033	0.164	0.839
Cluster 4	0.132	0.143	0.355
Cons	2.180	0.792	0.006
Nb. of observations	70		
R-sq	0.8776		
Chi2	444.59		

In Table 3 we present the results of the second 3SLS equation, where attendance is regressed on park characteristics. First, an increase in the number of thrill rides or family rides increases attendance. This result does not hold statistically for kiddie rides, water rides and shows. This may be explained by the fact that, in most parks, kiddie rides is a complement to the other kinds of rides. As far as shows are concerned, they are never the core business of a park, but part of general supply, on which parks do not rely to capture new large market shares. Second, the level of quality is, as expected, significant and positively related to attendance. However, theming is not significant. This surprising result could be partially explained by the fact that the variable *quality* captures part of the variable *theme*. There also seems to be a reputational effect in the sense that older parks are associated to higher attendance. Not surprisingly, parks located in more crowded areas have higher attendance, which means that park location is an important determinant of success. Concerning countries fixed effects, all Nordic countries have higher attendance than Belgium. This is also the case for France, Spain and Italy.

3.4. Theme parks ranking. After controlling for the presence of outliers and down-weighting them, the previous regression helps us to generate a ranking of parks. In terms of entry prices, the ten most estimated overpriced and underpriced parks to go to are reported in Table 4. Obviously, this is keeping in mind that still 10% of prices cannot be explained by our regression analysis and that the relative over-price could be simply due by a difference in a characteristic that is not available in the analysis[2].

Table 4. Overpriced (left) and underpriced (right)

Rank	Overpriced	Rank	Undepriced
1.	Legoland Deutschland (DE)	1.	Wild und Freizeitpark Klotten (DE)
2.	Fort Fun (DE)	2.	Bonbonland (DK)
3.	Alton Towers (UK)	3.	Skyline Park (DE)
4.	Holiday Park (DE)	4.	Fraispertuis City (FR)
5.	Thorpe Park (UK)	5.	Flamingoland (UK)
6.	Rastiland (DE)	6.	Centropark (DE)
7.	Walibi Rhone Alpes (FR)	7.	Parque Warner (SP)
8.	Karolinel und-Tivoliland (DK)	8.	Oakwood (UK)
9.	Gardaland (IT)	9.	Bobbejaanland (BE)
10.	Schwaben Park (DE)	10.	Sarkanniemi (FI)

Such a ranking obviously has to be read carefully and jointly with Figure 1. Legoland Deutschland is, given its characteristics, overpriced by 65% and Fort Fun by more than 50%. Disneyland Paris is in line with its characteristics while Walt Disney Studios is overpriced by 8.62%. Conversely, Klotten in Germany and Bonbonland in Denmark are underpriced by 30%. For some parks such as Frais-

[1] Note that discounts lead in fact to the same spending per person. Special offers and rebates are often limited bundles set to boost entries in off-season.

[2] The full ranking is reported in Appendix B.

pertuis City or Klotten, this could be explained by their far location from any large cities.

Conclusion and future development

In this paper, we study the theme parks market. We started by constructing a typology of parks (using the standard clustering methods) using the following variables: theme, quality, targeted markets and size. This allowed us to create groups of similar European parks. We find that there is no specific correlation between theme and quality since high quality but low themed parks, and vice-versa, exist.

In a second step, we looked for the determinants of entry prices and attendance of theme parks. A first fact that emerges is that both variables are highly and positively correlated. A proposed explanation is that consumers choose a park depending on the attractions one can find, and that prices are set by owners depending on the investment made. However, investments made in theming are surprisingly not significant, neither on prices nor attendance. Studying both variables separately, we find that there is no outlier in terms of attendance, meaning that for all parks, attendance is in line with their characteristics. In terms of prices, an analysis of the residuals shows that some parks are overpriced while others are underpriced. In particular, parks such as Legoland Deutschland turn out to be proportionally more expensive, by up to 65%, than the others even when controlling for their individual characteristics.

An interesting extension of the paper could be to suggest that the European market is very different from the US. Indeed the two directions that prevail in the US market – cloning parks and creating holiday areas – cannot be found in Europe. However, a thorough comparison should be supported by extending our dataset to American theme parks. Besides, an additional dimension could be added thanks to a spatial regression model, taking into account the matrix of distances between parks to control for any substitution or complementarity effect between parks.

Acknowledgments

We are grateful to Marjorie Gassner, François Koulischer and Christel Van Ranst for their careful reading and comments.

References

1. TEA/ERA (2006, 2007, 2008). Theme Park Attendance Report.
2. Adams, Judith (1991). The American Amusement Park Industry: a History of Technology and Thrills. Boston: Twayne Publishers.
3. Alter, Judy (1997). Amusement Parks. New York: Franklin Watts.
4. Bigné J. et al. (2005). The theme park experience: an analysis of pleasure, arousal and satisfaction, *Tourism Management*, Vol. 26 (6), pp. 833-844.
5. Anton-Clavé S. (2004). The global theme park industry. Cabi, p. 480.
6. Fodness D. and L. Milner (1992). A perceptual mapping approach to theme park visitor segmentation, *Tourism Management*, Vol. 13 (1), pp. 95-101.
7. McClung G.W. (1991). Theme park selection: factors influencing attendance, *Tourism Management*, Vol. 12 (2), pp. 132-140.
8. Moutinho, L. (1988). Amusement park visitor behaviour – Scottish attitudes, *Tourism Management*, Vol. 9 (4), pp. 291-300.
9. Swarbrooke J. (2002). “Leading theme parks in the USA”, Development and Management of Visitor Attractions (Second Edition), pp. 361-363.
10. Samuelson, D. and W. Yegoiants (2001). The American Amusement Park. St. Paul, MN: MBI Publishing Company.
11. Verardi, V. and C. Croux (2009). Robust regression in Stata, *Stata Journal*, StataCorp LP, Vol. 9 (3), pp. 439-453.
12. Verardi, V. and C. Dehon (2010). Multivariate Outlier Identification in Stata, *Stata Journal*, StataCorp LP, Vol. 10 (2), pp. 259-266.
13. Wanhill S. (2008). Chapter 4 – Economic Aspects of Developing Theme Parks, *Managing Visitor Attractions* (Second Edition), pp. 59-79.
14. Weiermair K. and Christine Mathies (2004). The tourism and leisure industry: shaping the future. Edition: illustrated, Haworth Press, p. 357.

Appendix A

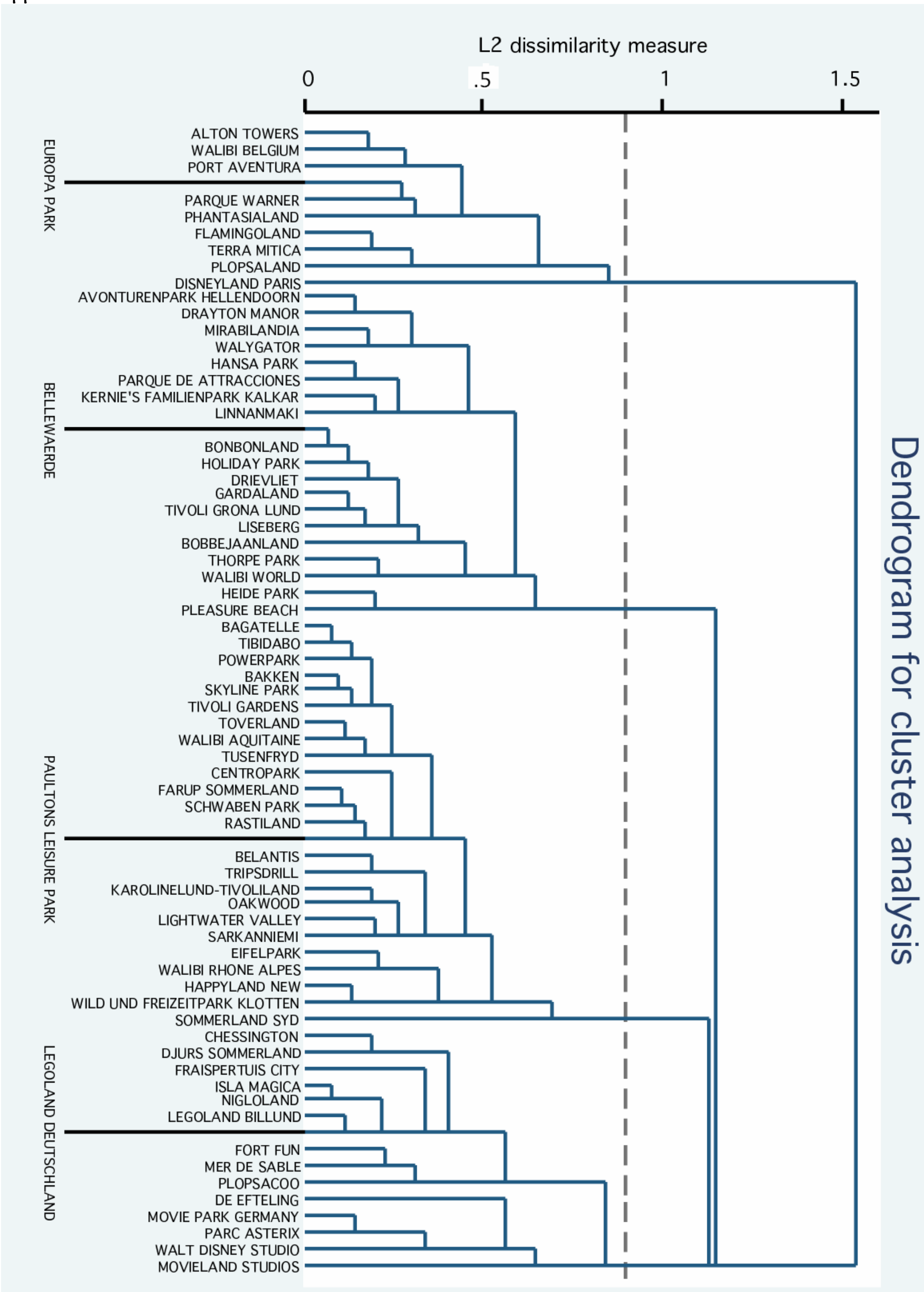

Fig. 1. Dendrogram for cluster analysis (TEA/ERA, 2006, 2007, 2008)

Appendix B

Table 1. Ranking in terms of price

Name	Country	Overprice
Legoland Deutschland	DE	64,74%
Fort Fun	DE	56,00%
Alton Towers	UK	26,70%
Holiday Park	DE	15,65%
Thorpe Park	UK	15,24%
Rastiland	DE	14,80%
Walibi Rhone Alpes	FR	13,27%
Karolinelund-Tivoliland	DK	13,06%
Gardaland	IT	12,08%
Schwaben Park	DE	10,89%
Plopsacoo	BE	10,76%
Tivoli Gardens	DK	9,80%
Belantis	DE	9,23%
Linnanmaki	FI	9,19%
Walt Disney Studio	FR	8,62%
Port Aventura	ES	7,90%
Kernie's Familienpark Kalkar	NL	7,53%
Plopsaland	BE	7,38%
Legoland Billund	DK	6,30%
Walibi Aquitaine	FR	5,78%
Liseberg	SE	5,61%
Powerpark	FI	5,37%
Avonturenpark Hellendoorn	NL	4,79%
Chessington	UK	4,67%
Tripsdrill	DE	4,37%
Lightwater Valley	UK	3,47%
Parc Asterix	FR	2,98%
Farup Sommerland	DK	2,79%
Movie Park Germany	DE	1,62%
Bellewaerde	BE	1,52%
Disneyland Paris	FR	1,19%
Terra Mitica	ES	0,75%
Mirabilandia	IT	0,46%
Tusenfryd	NO	-0,01%
Happyland New	CH	-0,03%
Parque de Attracciones	ES	-0,23%
Toverland	NL	-0,29%
Walygator	FR	-0,94%
Paultons Leisure Park	UK	-1,25%
Walibi World	NL	-1,66%
Djurs Sommerland	DK	-2,23%
Pleasure Beach	UK	-2,37%
Heide Park	DE	-2,51%
Tibidabo	ES	-2,65%
De Efteling	NL	-2,95%
Bagatelle	FR	-3,16%
Walibi Belgium	BE	-4,36%
Sommerland syd	DK	-4,46%
Europa Park	DE	-4,74%
Tivoli Grona Lund	SE	-5,34%
Isla Magica	ES	-5,36%
Hansa Park	DE	-5,67%

Table 1 (cont.). Ranking in terms of price

Name	Country	Overprice
Drayton Manor	UK	-6,33%
Drievliet	NL	-6,84%
Eifelpark	DE	-8,92%
Bakken	DK	-10,84%
Movieland Studios	IT	-11,22%
Phantasialand	DE	-11,68%
Mer de Sable	FR	-11,77%
Nigloland	FR	-13,00%
Sarkanniemi	FI	-13,12%
Bobbejaanland	BE	-13,47%
Oakwood	UK	-14,02%
Parque Warner	ES	-15,40%
Centropark	DE	-17,00%
Flamingoland	UK	-18,67%
Fraispertuis City	FR	-22,34%
Skyline Park	DE	-30,41%
Bonbonland	DK	-31,76%
Wild und Freizeitpark Klotten	DE	-33,77%