Regional bias when benchmarking services using customer satisfaction scores

Andrew Brint[a] & John Fry[b]

[a] *Sheffield University Management School, 1 Conduit Road, Sheffield, S10 1FL, UK*

a.brint@sheffield.ac.uk

[b] *School of Computing Mathematics and Digital Technology, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester M1 5GD, UK*

j.fry@mmu.ac.uk

Regional bias when benchmarking services using customer satisfaction scores

Regional monopoly service organisations such as electricity, gas and water distributors, health trusts, public transport, and local government are subject to regulatory oversight. A common element in this is benchmarking an organisation against similar organisations based in different regions. Customer satisfaction is often an important part of this competitive benchmarking. However, if people from different regions give a different average satisfaction score for the same experience, then this disadvantages some companies. Therefore, regional satisfaction was investigated in an environment where differences in customer service levels are controlled for. The average online satisfaction ratings people from different regions of the UK gave to the same overseas holiday hotels were investigated. The 24,154 ratings were analysed using linear mixed effects and ordinal models. The average ratings given by people from the London region were significantly lower than those from elsewhere. Regional correction factors are developed and applied to published satisfaction ratings for electricity distributors. The adjustment was sufficient to move the London distributor from the penalty category to a borderline position. Hence, customer satisfaction ratings should be used cautiously when benchmarking regional organisations. This investigation of the potential for regional bias contributes to the large literature on customer satisfaction and behavioural intentions.

Keywords: competitive benchmarking; monopoly services; regional satisfaction; customer satisfaction; behavioural intentions.

1. Introduction

Regional monopoly service providers such as energy and water distributors, health trusts, and local authorities, need to be monitored to ensure that they are delivering a satisfactory level of performance. Competitive benchmarking is a natural and important way to do this (Pollitt 2005, Trompet et al. 2013, Le Lannier & Porcher 2014). As users' satisfaction is an important measure of the performance of public service providers, customer feedback ratings are key elements in many benchmarking exercises

(Magd & Curry 2003, Guglielmetti Mugion & Musella 2013). For example, the UK's Office of Gas and Electricity Markets (Ofgem 2012) uses satisfaction surveys as a formal element of the assessment of electricity and gas distributors. However, when feedback ratings are used to compare regional monopolies such as electricity network owners, there is the potential for regional bias in the average satisfaction score, i.e. some regions giving lower average scores than others. This would leave the monopolies in some regions at a disadvantage when compared with those in other regions. Ofgem (2012) details the significant financial penalties that could be applied to distributors whose satisfaction scores are too far away from the mean satisfaction score. The investigation was motivated by the concern of some regional electricity and gas distributors in the UK that regional bias was affecting the imposition of these penalties. However, they could not provide evidence to Ofgem of a regional effect. Consequently, this paper provides a statistical investigation of the following questions:

Is there a regional bias in customer satisfaction surveys?

If there is a regional bias, then how can it be corrected for?

Understanding these issues is important so that sound decisions can be made. A large academic literature discusses the relationship between customer satisfaction and service quality (see e.g. Cronin and Taylor, 1992). We study this in an environment where we can control for service quality.

The investigation analysed the regional breakdown of the online ratings of overseas (i.e. non-UK) holiday hotels posted by UK residents on TripAdvisor. Ratings given by reviewers from the London region were significantly lower than those from other regions, leading to the London based utilities being disadvantaged. Regional correction factors are devised and applied to published satisfaction ratings for UK

electricity distributors. The literature suggests that regional personality differences are not just a UK phenomenon, and so there are very likely to be similar problems with applying competitive benchmarking using customer feedback ratings in other countries. The approach described in this paper can easily be implemented in these cases.

The layout of this paper is as follows. Section 2 very briefly summarizes literature that suggests that personality measurements differ between regions, before identifying the very limited amount of work that has been published on how customer satisfaction ratings vary between regions within a country. Behavioural aspects of customer satisfaction are also reviewed. Section 3 describes the methodology and the data that the study used. The linear mixed effects model analysis and its results are presented in Section 4. Elementary robustness checks are carried out in Section 5. Section 6 investigates whether socio-economic factors can explain the regional differences found in Section 4. Finally, Section 7 discusses the assumptions behind the analysis and the implications of the results, before Section 8 concludes and considers the opportunities for further work.

2. Related literature

Differences in average personalities (Allik 2012) and customer satisfaction levels (Qi et al. 2012) exist between developed countries. Also differences in survey response patterns can be expected between cultures (Mattila 1999, Hofstede 2005, Grün 2007, Reynolds & Smith 2010, Yayla-Küllü et al. 2015). Consequently, Trompet et al. (2013, page 3) notes that "there are socio-political, structural and cultural differences that exist between cities in different countries, which can lead to a 'cultural bias' when comparing satisfaction directly". However, while perceptions of regional differences in attitudes have a long history, one might expect that the unprecedented mobility of modern

populations, coupled with the penetration of national and international media, may have eliminated the differences and led to an increasingly homogeneous culture (Plaut et al. 2002). This is not the case though as differences in personality between regions within countries are still being reported (Plaut et al. 2002, Rentfrow et al. 2015). Hence it is possible that some regions within a country could be more reserved with their praise (and maybe their criticism) than others when filling out customer feedback surveys. If this is the case, then the use of customer satisfaction scores could lead to an inappropriate ranking of companies operating in different regions.

Before considering the work that has been reported on regional personality differences and regional satisfaction levels, the behavioural aspects of customer satisfaction are briefly reviewed.

2.1 Behavioural aspects of customer satisfaction

Customer satisfaction is an important part of quality management for the service industry (see e.g. Kyriakopoulos, 2011 for a review, and Kursunluoglu, 2014 for how it is defined and measured). It is often associated with a plethora of desirable customer post-purchase intentions (González et al., 2007; Minh et al., 2015). Customer satisfaction is inherently subjective in a services rather than a manufacturing setting (Koc, 2006). Hence, some authors define customer satisfaction as an emotional reaction to the gap between expectations of a product or service and the actual customer experience (Oliver, 1981). Perceptions of overall quality – as distinct from the actual level of service delivered – are also important within highly competitive industries (Hu et al., 2009). Therefore, the relationship between customer satisfaction and service quality has been extensively studied (see e.g. Parasuram et al., 1988; Cronin and Taylor, 1992; Clemes et al., 2011). However, it is important to measure service quality in as

reliable and scientific a way as possible – aside from subjective elements of customer perception (Xie, 2011). Thus, we study customer satisfaction in an environment where we can control for service quality (see Section 3).

Allied to the above, our study of regional bias adds to a range of behavioural aspects of customer satisfaction surveys that have been discussed in the literature. Nankoo et al. (2017) find a positive relationship between service quality and the subjective consumption emotions experienced when purchasing products or services. Customer satisfaction has also been variously associated with intangible rather than tangible product features (Minh et al., 2015), employee satisfaction (Pantouvakis and Bouranta, 2013) and company image (Silvestri et al., 2017). There may also be strict limits as to the extent to which consumers can rationally evaluate additional technical product features (Thompson et al., 2005).

2.2 Regional personality differences

Plaut et al. (2002) found that differences in personality profiles still exist between regions in modern day USA. For example, the New England personality profile was less assertive and dominant than the Mountain region profile. Similarly, surveys have shown that regional differences in both happiness and attitude also exist in the UK. For example, Rentfrow et al. (2015) mapped the Big Five Personality traits of extroversion, agreeableness, conscientiousness, neuroticism and openness against UK geographical districts using the almost 400,000 responses to an online quiz. It was found that there were distinct geographical patterns in the personality traits. Out of the Big Five Personality traits, agreeableness would seem to be the one most likely to have an impact on customer satisfaction ratings. Rentfrow et al. (2015) found that London was the region with the lowest level of agreeableness. This may have important regulatory

implications for regional monopolies if low levels of agreeableness correlate with a tendency to give lower ratings.

The regional differences in happiness and satisfaction with life are sufficiently important for the UK government to regularly monitor them (ONS 2014). London is again the outlier with markedly lower happiness and satisfaction ratings.

2.3 Regional satisfaction surveys

There have been very few reported studies looking at and quantifying how satisfaction with the same quality of product or service varies between regions within a country. Huang et al. (2011) investigated how satisfaction with products differed between developed and non-developed regions of China. It was found that regions with higher levels of economic development were associated with the more critical rating of products. This fits in with the finding of Grigoroudis et al. (2008) that satisfaction levels are linked to a country's economic performance.

Zhang et al. (2013) studied how different regional factors such as economic development could affect customer satisfaction in China – this time focusing upon the rating of a chain of restaurants. However, in this case, differing regional tastes and differences between individual restaurants may mean that the products involved are not directly comparable. (Dermanoff & Eklöf (2001) note a number of issues that make comparing satisfaction levels of different services at different places and times difficult.)

3. Methodology

The online ratings that UK customers have given to overseas holiday hotels on TripAdvisor were analysed to see whether there was an overall difference in scores between regions. Besides being one of the most popular hotel rating websites

(Blomberg-Nygard & Anderson 2016, Yu et al. 2017), TripAdvisor has been the focus of a number of academic analyses of rating behaviour (Banerjee & Chua 2016), particularly in relating the score given to the accompanying written review (Mankad et al. 2016, Guo et al. 2017, Liu et al. 2017) and how the written reviews mediate the trustworthiness of the ratings (Gavilan et al. 2018). The focus of this paper differs from these analyses as it investigates regional differences in customer rating levels rather than factors that cause one hotel to be preferred to another. Allik (2012) notes that the 'variance produced by cross-cultural differences is approximately nine times smaller than what is produced by interindividual variance within each country.' Hence the variance of regional differences is likely to be low compared with between individual variance, and so a large data set is likely to be needed to quantify any regional difference. Therefore, ratings of overseas holiday hotels were chosen as:

- Individual hotels often have very large numbers of ratings.
- There is no reason to believe that a particular hotel would appeal more to one region than to another.

European hotels used by two of the UK's leading package tour operators, First Choice / TUI and Thomas Cook, were identified from the companies' printed family holiday brochures. A random selection was made from these hotels subject to them having more than 200 ratings with the majority of these ratings being from UK reviewers and subject to no more than one hotel from each resort and including as many countries as possible. Overall 51 hotels were selected. Only reviews submitted in the preceding 24 months were considered. The data noted from each review was the rating (in the range 1 to 5) and the region of the UK the reviewer was from. The latter was carried out by identifying their specified town or city with a county, and then grouping

the counties into regions. Altogether 24,154 ratings were analysed. Descriptive statistics for the data are shown in Tables 1 and 2.

A potential problem with any survey is that the participants may not be representative of the whole population (Anderson, 1998). However, an important advantage offered by the hotel data set is that it is centred on the online rating of an experience, as this matches the situation of interest, i.e. the online rating of customer satisfaction. Another concern is the impact of inauthentic reviews. However, Blomberg-Nygard & Anderson (2016) argues that the huge number of reviews on TripAdvisor and the filters used by TripAdvisor to remove suspect reviews, means that the impact of inauthentic entries is minimized. Additionally, the effects of demographic and economic differences (Van de Vijver & Leung 1997) are reduced by the grouping of reviewers on the basis of having selected a particular type and cost of holiday. The issue of the suitability of the data set is considered further in the discussion section.

The data was categorized into the following regions:

- NW   North West England

- NE   North East England and Yorkshire

- WM   West Midlands

- EME East Midlands and East England

- L   London

- SE   South East England

- SW   South West England

- SC   Scotland

- W   Wales

This regional division of England follows the standard UK government model (ONS 2016d) apart from the combining of two neighbouring government regions into one region for

- North East England and Yorkshire & Humberside to create the NE region
- East Midlands and East to create the EME region

This merging was done so as to ensure that each English region had at least 2,000 responses.

As shown in equation (1) below, the data has a hierarchical structure that enables us to isolate a regional effect once differences in individual hotels are accounted for. The first level of correlation occurs between ratings from the same hotel – corresponding to the same customer experience. A secondary correlation exists between ratings for the same region – though the effect is reduced compared to ratings from the same hotel. Finally, the data will also show idiosyncratic subject-level variation reflecting differences between individual survey respondents.

4. Analysis

The overall customer satisfaction rating is modelled as a linear mixed effects model. The reasons for approaching the analysis in this way are threefold. Firstly, this allows any identified regional effect to be precisely quantified. Secondly, this approach allows for a relatively straightforward way to adjust for a regional effect in regulatory applications should such a bias be found. Thirdly, in setting up the model in this way an independent check of the findings with ordinal regression analyses can be carried out (see Section 5: Ordinal analysis of ratings).

Let

$$O_{ijk} = \mu + h_i + r_j + \varepsilon_{ijk} \tag{1}$$

where O_{ijk} is the overall rating from the kth review from the set of hotel i's region j reviews, μ is the underlying mean, h_i is the random effect of hotel i, r_j is the fixed effect of region j and the ε_{ijk} are independent normally distributed errors. Region r_j is modelled as a fixed effect as the value of the regional effect for each region is the key object of interest. Hotel h_i is modelled as a random effect as the data consists of repeated measurements for each hotel. Modelling the data in this way thus allows differences between individual hotels to be accounted for before testing for a regional effect.

Table 3 gives the results of fitting the mixed linear model of equation (1). The likelihood ratio test gave a χ^2 value of 45.2 on 8 degrees of freedom (p=0.0000). Therefore, the null hypothesis of no regional fixed effect can be rejected.

The estimated coefficient for the London region is negative with a value of −0.116 and is statistically significant (p=0.000). This suggests that after account is taken of the different hotels, the London region is more likely to give lower ratings. There is some evidence (p=0.040) that Scotland may give higher ratings than other regions. However, this finding is not robust to a series of robustness checks performed in Section 5. There is no evidence for differences between the other UK regions. As a simple corollary to Table 3, a mean centring correction (Fischer 2004) for regional bias can be obtained by reversing the signs of the estimated fixed effects terms. The suggested corrections based on the current data set are shown in Table 4.

5. Ordinal analyses of ratings

Customer satisfaction ratings are predominantly handled in practice as if they are interval data so allowing their means and variances to be calculated and compared.

However, it can be argued that they should be treated as ordinal data. Therefore, rating differences between the regions were also investigated when the ratings were treated as being ordinal rather than interval data.

5.1 Regional personality differences

An ordinal logistic regression model with the assumption of proportional odds (Kleinbaum & Klein 2002) where the probability of a review of hotel i in region j giving a rating of m is

$$\frac{1}{1+e^{-\mu_m-h_i-r_j}} \tag{2}$$

was fitted to the data. The results are given in Table 5. The only significant p-value is for the London region (with a value of 0.007) providing an indication that reviewers from the London region give less reviews with higher numbers of stars, e.g. an estimated 38% being 5 star compared with 39% from the North East & Yorkshire region while all the other regions gave 40% or more.

5.2 Net balance of positive and negative reviews

Ratings were categorized as low (1 or 2 stars), medium (3 stars) and high (4 or 5 stars) as in Valdivia et al. (2017) and logistic generalized linear mixed models were fitted to each of these three cases in turn to see if there were regional differences in the incidence of low, medium and high scores. The probability of a review of hotel i in region j falling into the relevant category (high, low or medium) is given by

$$\frac{1}{1+e^{-\mu-h_i-r_j}}$$

i.e. the same as equation (2) apart from the omission of the subscript m on μ as rather

than having 5 possibilities for the rating, the review is now either in the category or not in the category. The results are given in Table 6. The London region has a significant effect on both low ratings (p-value = 0.0012) and high ratings (p-value = 0.0003), but there were no other significant regional effects for low or high scores. No regions had significant regional effects for medium scores. Thus, it appears that the London region is associated with both relatively lower numbers of higher ratings and higher numbers of lower ratings.

6. Socio-economic factors

Huang et al. (2011) and Grigoroudis et al. (2008) have linked reported satisfaction levels with respectively regional and economic performance. Therefore, the approach in Rentfrow et al. (2015) was followed and the effects of key demographic, political, economic, social and health indicators were examined. The indicators considered were:

Demographic indicators – Population statistics for each area were obtained (ONS 2016c) together with the proportion of males in each region and the median age of residents in each region (ONS 2016c).

Political indicators – The share of the vote for the three main political parties (Conservative, Labour and Liberal Democrat) in the 2015 general election were used (Hawkins et al., 2015).

Economic indicators – Wealth and human capital data were used as economic indicators. For wealth, the estimates of median annual income for 2014 were obtained from the Office of National Statistics Annual Survey of Hours and Earnings (ONS 2016a). Data was also gathered on the percentage of the working-age population with no qualifications in 2012 obtained from the Annual Population Survey (ONS 2016b).

Social indicators – Information was gathered about cultural diversity, marital status and crime. For an indicator of cultural diversity, the information used was the international migration (inflow) (ONS 2016c). Following a broadly similar approach to Rentfrow et al. (2015), the proportion of the population either married or cohabiting was used as an indicator for the degree of social stability in the region (ONS 2016c). As an indicator of criminal activity, the recorded offences per 1,000 of the population were used (ONS 2016c).

Health indicators – Regional information on the male and female life expectancy at birth from 2009-2011 (ONS 2016d) was included.

The indicators described above were included in the linear mixed effect model. Let d_j denote a demographic variable corresponding to region j. We tested the null hypothesis of no regional effect

$$O_{ijk} = \mu + h_i + \alpha\, d_j + \varepsilon_{ijk}$$

against the alternative model

$$O_{ijk} = \mu + h_i + \alpha\, d_j + r_j + \varepsilon_{ijk}$$

where r_j denotes the fixed regional effect. A likelihood ratio test was used to test the null hypothesis of no regional effect once the demographic variables had been included in to the model. The results are given in Table 7. In all the cases the p-value is below 0.005, and so the identified regional affect still remains highly significant even when these additional explanatory variables are included in the model.

7. Discussion

In this section we show that the problem of detecting regional bias in customer satisfaction surveys may have practical significance for business with an application to

14

the benchmarking of regional monopoly service organisations. Within this context customer satisfaction scores are important as they are often used as the basis for regulatory fines to penalise perceived poor performance. The mixed effects model shown in equation (1) can be used to estimate the size of the regional bias (Table 3) and propose a correction factor (Table 4). The effect of applying the correction adjustments of Table 4 to the regional satisfaction scores of the UK's electricity customers (Ofgem 2013) is shown in Table 8. The regions correspond with the electricity company boundaries. Column 2 gives the average satisfaction for each company, and Column 3 is the column Ofgem uses to assess the companies. Column 4 is the number of standard deviations away from the overall mean a company's satisfaction score is after applying the adjustments specified in Table 4. A suitable adjustment was made when a company's area in Table 8 was in more than one of the regions of Section 3. The main change between Columns 3 and 4 is the movement of the London company from 2.14 standard deviations below the mean to 1.77 standard deviations below the mean. As the critical value is often taken to be 1.75 (Ofgem 2012) or 2.00 standard deviations below the mean, changes of this nature can have significant financial implications.

The main assumption with the analysis is that regional differences in the online satisfaction ratings of overseas holiday hotels, is representative of a more general regional difference in the assignment of satisfaction ratings. There are two elements to this assumption:

- Regional 'biases' in online reviewers reflect 'biases' in the regional populations as a whole.

- Hotel ratings are suitably representative of a wide range of different products and services.

Firstly, since the customer rating of organisations is increasingly being done through online surveys, any regional biases by online reviewers will increasingly impact on the rating of regional organisations even if there is no bias in the population as a whole. Although there is no reason to believe that a regional bias in online ratings would not also be reflected in a regional bias in people who do not submit online ratings, this is not a necessary assumption behind the analysis. Hence the suggested regional adjustment to ratings is likely to be valid irrespective of potential concerns that those who post online ratings constitute a self-selecting sample.

Secondly, both academic studies of personality (Rentfrow et al. 2015) and the government measures of happiness (ONS 2014) give corroborating evidence of differences between London and the rest of the UK. This suggests that the results from analysing hotel ratings would continue to be the case when rating products and services in general. A significant proportion of the UK population has overseas holidays, and sites such as TripAdvisor have large numbers of reviews, and so hotel ratings reflect the views of a broad swathe of the population. However, it would be valuable to perform further analyses similar to the present one but with other products, so as to confirm the effect.

8. Conclusions

The relationship between customer satisfaction and service quality is much studied (see e.g. Parasuram et al., 1988; Cronin and Taylor, 1992; Clemes et al., 2011). In this paper we show how to quantify behavioural intentions once an adjustment has been made for the underlying service quality. The importance of our contribution is threefold. Firstly, regional aspects of customer satisfaction, personality differences and survey bias are under-explored in the literature. Secondly, such problems can be shown to have

practical significance with respect to the benchmarking of regional monopolies (especially utilities) in the UK. Thirdly, our study, and the presence of a regional effect, adds to a documented list of behavioural factors that affect customer satisfaction including consumption emotions (Nankoo et al., 2017), intangible rather than tangible product features (Minh et al., 2015), employee satisfaction (Pantouvakis and Bouranta, 2013), company image (Silvestri et al., 2017) and consumer irrationality (Thompson et al., 2005).

The weight of evidence suggests that there is a modest, but clearly discernible and statistically significant, regional effect in customer satisfaction ratings. Reviewers from London give lower average ratings than reviewers from other regions in the UK. These findings are in accord with evidence from both academic (Rentfrow et al. 2015) and government (ONS 2014) studies. The literature suggests that it is unlikely that the UK is the only country with this regional effect. The approach described in this paper can be used to quantify the size of these effects in other countries, and so to correct for them. For example, when benchmarking bus providers in Brazil, Lindau et al. (2017) note that "significant differences among cities may infer … that the evaluation in perceptions is different for each city" (page 6). Applying the methodology of this paper to Brazilian cities would improve the validity of the inter-city comparisons.

Our study is inevitably limited by the number of products (hotels) studied and our restriction to the nine listed UK regions given in Section 3. Extensions to other products and regions/countries are possible and would mirror recent novel applications of related big-data sources (see e.g. Banerjee and Chua, 2016; Guo et al., 2017; Mankad et al., 2016). One direction could be to consider the section scores, e.g. "Cleanliness", given by reviewers in TripAdvisor as these are separate from the overall score

(Martin–Fuentes et al. 2018; Diaz and Rodriguez 2018), and so could provide an additional source of data.

A further limitation is the limited demographic information available on survey respondents. Future work could incorporate additional variables such as age, gender, economic profile and occupation of those surveyed. However, this may come at the cost of reduced sample sizes and less reliable estimates of the size of the regional effect.

References

Allik, J. 2012. National differences in personality. *Personality and Individual Differences* **53**(2) 114-117.

Anderson, E.W. 1998. Customer satisfaction and word of mouth. *Journal of Service Research* **1**(1) 5-17.

Banerjee, S., A.Y.K. Chua. 2016. In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management* **53** 125-131.

Blomberg-Nygard, A., C.K. Anderson. 2016. United Nations World Tourism Organization study on online guest reviews and hotel classification system: An integrated approach. *Service Science* **8**(2) 139-151.

Clemes, M. D., C. Gan, M. Ren. 2011. Synthesizing the effects of service quality, value and customer satisfaction on behavioural intentions in the model industry: An empirical analysis. *Journal of Hospitality and Tourism Research* **35**(4) 530-568.

Cronin, J., S. Taylor. 1992. Measuring service quality: A re-examination and extension. *Journal of Marketing* **56**(3) 55-88.

Dermanoff, V., J. Eklöf. 2001. Using aggregate Customer Satisfaction Index: Challenges and problems of comparison with special reference to Russia. *Total Quality Management* **12**(7-8) 1054-1063.

Díaz, M. R., T. F. E. Rodríguez. 2018. Determining the reliability and validity of online reputation databases for lodging: Booking.com, TripAdvisor, and HolidayCheck. *Journal of Vacation Marketing* **24**(3) 261-274.

Fischer, R. 2004. Standardization to account for cross-cultural response bias: a classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology* **35**(3) 263-282.

Gavilan, D., AM. Avello, G. Martinez-Navarro. 2018. The influence of online ratings on hotel booking consideration. *Tourism Management* **66** 53-61.

González, M. E. A., L. Comesaña, J. A. F. Brea. 2007. Assessing tourist behavioural intentions through perceived service quality and customer satisfaction. *Journal of Business Research* **60**(2) 153-160.

Grigoroudis, E., G. Nikolopoulou, C. Zopounidis. 2008. Customer satisfaction barometers and economic development: An explorative ordinal regression analysis. *Total Quality Management & Business Excellence* **19**(5) 441-460.

Grün, S.D.B. 2007. Cross-cultural differences in survey response patterns. *International Marketing Review* **24**(2) 127-143.

Guglielmetti Mugion, R., F. Musella. 2013. Customer satisfaction and statistical techniques for the implementation of benchmarking in the public sector. *Total Quality Management & Business Excellence* **24**(5-6) 619-640.

Guo, Y., S.J. Barnes, Q. Jia. 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management* **59** 467-483.

Hawkins, O, R. Keen, D.G. Clayton. 2015. General election 2015
House of Commons Briefing Paper **CBP7186**
http://researchbriefings.files.parliament.uk/documents/CBP-7186/CBP-7186.pdf

Hofstede, G.H. 2005. *Cultures and Organizations: Software of the Mind.* Updated Edition, McGraw-Hill, London, UK.

Hu, H-H., J. Kandampully, T. D. Juwaheer. 2009. Relationships and impacts of service quality, perceived value, customer satisfaction and image: an empirical study. *The Service Industries Journal* **29**(2) 111-125.

Huang, J., X. Wang, R. Chen. 2011. Regional differences in customer satisfaction in China. *Social Behavior and Personality* **39**(10) 1403-1412.

Kleinbaum, D.G., M. Klein. 2002. *Logistic regression.* 2nd Edition, Springer-Verlag, New York, NY.

Koc, E. 2006. Total quality management and business excellence in services: the implications of all-inclusive pricing system on internal and external customer satisfaction in the Turkish tourism market. *Total Quality Management* **17**(7) 857-877.

Kursunluoglu, E. 2014. Shopping centre customer service: creating customer satisfaction and loyalty. *Marketing Intelligence and Planning* **32**(4) 528-548.

Kyriakopoulos, G. L. 2011. The role of quality management for effective implementation of customer satisfaction, customer consultation and self-assessment, within service quality schemes: A review. *African Journal of Business Management* **5**(12) 4901-4915.

Le Lannier, A., S. Porcher. 2014. Efficiency in the public and private French water utilities: Prospects for benchmarking. *Applied Economics* **46**(5) 556-572.

Lindau, L. A., M. M. Barcelos, M. B. B. da Costa, C. S. ten Caten, C. A. M. DaSilva, B. M. Pereira. 2017. Benchmarking focused on the satisfaction of bus transit users. *Transportation Research Board 96ᵗʰ Annual Meeting Compendium of Papers* (available at https://trid.trb.org/view/1439239)

Liu, Y., T. Teichert, M. Rossi, H. Li, F. Hu. 2017. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management* **59** 554-563.

Magd, H., A. Curry. 2003. Benchmarking: achieving best value in public-sector organisations. *Benchmarking: An International Journal* **10**(3) 261-286.

Mankad, S., H.S. Han, J. Goh, S. Gavirneni. 2016. Understanding online hotel reviews through automated text analysis. *Service Science* **8**(2) 124-138.

Martin-Fuentes, E., C. Mateu, C. Fernandez. 2018. Does verifying uses influence rankings? Analyzing booking.com and tripadvisor. *Tourism Analysis* **23**(1) 1-15.

Mattila, A. A. 1999. The role of culture in the service evaluation process. *Journal of Service Research* **1**(3) 250-261.

Minh, N. H., N. T. Ha, P. C. Anh, Y. Matsui. 2015. Service quality and customer satisfaction: A case study of hotel industry in Vietnam. *Asian Social Science* **11**(10) 73-85.

Nankoo, R., V. Teeroovengadum, P. Thomas, L. Leonard. 2017. Integrating service quality as a second-order factor in a customer satisfaction and loyalty model. *International Journal of Contemporary Hospitality Management* **29**(12) 2978-3005.

Ofgem. 2012. Decision on the customer satisfaction survey incentive rate term in Part D of electricity distribution Special Licence Condition CRC8 https://www.ofgem.gov.uk/ofgem-publications/46559/incentive-rate-decision-290312.pdf

Ofgem. 2013. Electricity Distribution Network Operator (DNO) – Customer Satisfaction Survey Results Q1 2012/2013 https://www.ofgem.gov.uk/publications-and-updates/customer-satisfaction-survey-results-2012-13-quarter-1-overview

Oliver, R. L. 1981. Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing* **57** 25-48.

ONS. 2014. 'National Well-being measures: September 2014' Office of National Statistics http://www.ons.gov.uk/ons/guide-method/user-guidance/well-being/index.html

ONS. 2016a. 'Annual survey of hours and earnings' Office of National Statistics https://www.gov.uk/government/statistics/annual-survey-of-hours-and-earnings-2013-revised-results

ONS. 2016b. 'Annual population survey' Office of National Statistics http://www.ons.gov.uk/ons/search/index.html?newquery=annual+population+survey

ONS. 2016c. 'Region and country profiles – directory of tables' Office of National Statistics http://www.ons.gov.uk/ons/rel/regional-trends/region-and-country-profiles/index.html

ONS. 2016d. 'Regions (Former GORs)' Office of National Statistics http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html

Pantouvakis, A., N. Bouranta. 2013. The interrelationship between service features, job satisfaction and customer satisfaction: Evidence from the transport sector. *The TQM Journal* **25**(2) 186-201.

Parasuram, A., V. Zheithaml, L. Berry. 1988. SERVQUAL: a multiple item scale for measuring customer perceptions of service quality. *Journal of Retailing* **64**(1) 12-40.

Plaut, V.C., H.R. Markus, M.E. Lachman. 2002. Place matters: Consensual features and regional variation in American well-being and self. *Journal of Personality and Social Psychology* **83**(1) 160-184.

Pollit, M. 2005. The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks. *Utilities Policy* **13**(4) 279-288.

Qi, J.-Y., Y.-P. Zhou, W.-J. Chen, Q.-X. Qu. 2012. Are customer satisfaction and customer loyalty drivers of customer lifetime value in mobile data services: a

comparative cross-country study. *Information Technology and Management* **13**(4) 281-296.

Rentfrow, P.J., M. Jokela, M.E. Lamb. 2015. Regional personality differences in Great Britain. *PLOS ONE* **10**(3) 1-20.

Reynolds, N., A. Smith. 2010. Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research* **13**(2) 230-243.

Silvestri, C., B. Aquilani, A. Ruggieri. 2017. Service quality and customer satisfaction in thermal tourism. *The TQM journal* **29**(1) 55-81.

Thompson, D. V., R. W. Hamilton, R. T. Rust. 2005. Feature fatigue: when product capabilities become too much of a good thing. *Journal of Marketing Research* **42**(4) 431-442.

Trompet, M., R. Parasram, R. J. Anderson. 2013. Benchmarking disaggregate customer satisfaction scores between bus operators in different cities and countries. *Transportation Research Record: Journal of the Transportation Research Board* **2351** 14-22.

Valdivia, A., M.V. Luzón, F. Herrera. 2017. Sentiment Analysis in TripAdvisor. *IEEE Intelligent Systems* **32**(4), 72-77.

Van de Vijver, F., K. Leung. 1997. Methods and data analysis for cross-cultural research. SAGE: London.

Xie, X. 2011. Service quality measurement from customer perception based on services science, management and engineering. *Systems Engineering Procedia* **1** 337-343.

Yayla-Küllü, H.M., P. Tansitpong, A. Gnanlet, C.M. McDermott, J.F. Durgee. 2015. Employees' national culture and service quality: An integrative review. *Service Science* **7**(1) 1-18.

Yu, Y., X. Li, T.-M. Jai. 2017. The impact of green experience on customer satisfaction: evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management* **29**(5) 1340-1361.

Zhang, Z., Z. Zhang, R. Law. 2013. Regional effects on customer satisfaction with restaurants. *International Journal of Contemporary Hospitality Management* **25**(5) 705-722.

Table 1:　　　Frequency of the different ratings in the 24,154 reviews.

	Number	Percentage
5 star	9,660	40%
4 star	8,211	34%
3 star	3,744	16%
2 star	1,492	6%
1 star	1,047	4%

Table 2:　　　Regional breakdown of the 24,154 reviews.

East Midlands and East England	EME	2,878
London	L	3,091
North East England and Yorkshire	NE	3,330
North West England	NW	3,525
Scotland	SC	2,048
South East England	SE	3,497
South West England	SW	1,999
Wales	W	1,419
West Midlands	WM	2,367

Table 3: Fitted linear mixed effects model for the dependence of rating upon region

Random effects	Std dev		
Hotel (intercept)	0.297		
Residual	1.030		
Fixed effects	Estimate	Std error	p-value
Intercept (EME)	3.97	0.0485	0.000
NW	0.00203	0.0263	0.936
NE	-0.00146	0.0267	0.960
WM	0.0200	0.0291	0.490
L	-0.116	0.0273	0.000
SE	-0.0132	0.0265	0.617
SW	0.0135	0.0306	0.660
SC	0.0625	0.0304	0.040
W	0.0126	0.0339	0.711

Table 4: Suggested adjustment for correcting the observed ratings for regional bias

Region	Adjustment to rating
NW	-0.002
NE	0.001
WM	-0.020
EME	0.000
L	0.116
SE	0.013
SW	-0.013
SC	-0.063
W	-0.013

Table 5: Ordinal logistic regression (equation 2) for the fraction of each number of stars awarded by reviewers from each region. West Midlands was the redundant parameter.

Fixed	r_j estimate	Std error	p-value
EME	-.030	.051	.558
L	-.134	.050	.007
NE	-.066	.049	.182
NW	-.020	.049	.679
SC	-.049	.055	.372
SE	.009	.049	.848
SW	.030	.056	.594
W	-.007	.062	.915
WM	---	---	---
Random	**Std dev**		
Hotel	.0005		

Table 6: Logistic generalized linear mixed models for the incidence of low, medium and high scores

Effects	Low ratings			Medium ratings			High ratings		
Random	Intercept			Residual			Intercept	Residual	
Std dev	0.639	0.989		0.378	0.996		0.544	0.997	
Fixed	Estimate	Std error	p-value	Estimate	Std error	p-value	Estimate	Std error	p-value
Intercept (EME)	-2.35	0.117	0.0000	-1.66	0.0782	0.0000	1.08	0.0932	0.0000
NW	0.124	0.0848	0.1430	-0.0961	0.0711	0.1767	0.00741	0.0602	0.9021
NE	0.0786	0.0859	0.3603	-0.0382	0.0711	0.5909	-0.00940	0.0607	0.8770
WM	-0.0244	0.0965	0.8005	0.00738	0.0774	0.9241	0.00546	0.0666	0.9346
L	0.281	0.0869	0.0012	0.112	0.0721	0.1220	-0.220	0.0616	0.0003
SE	0.0527	0.0876	0.5480	-0.0295	0.0713	0.6788	-0.00590	0.0609	0.9228
SW	0.0585	0.100	0.5586	-0.107	0.0834	0.1996	0.0468	0.0704	0.5060
SC	-0.0957	0.0988	0.3327	-0.0458	0.0806	0.5701	0.0851	0.0691	0.2185
W	-0.0332	0.113	0.7681	-0.0585	0.0915	0.5226	0.0572	0.0781	0.4637

Table 7: Maximum likelihood ratio tests of the null hypothesis of no regional effect once the additional demographic variables are included into the model

Model (fixed effects)	χ^2	p-value
Population & Region	36.0	0.0000
Gender & Region	22.0	0.0000
International migration & Region	8.0	0.0047
Age & Region	8.0	0.0047
% Conservative vote & Region	42.0	0.0000
% Labour vote & Region	36.0	0.0000
% Lib Dem vote & Region	44.0	0.0000
Income & Region	10.0	0.0016
% of workforce with no qualifications & Region	38.0	0.0000
% Married of cohabiting & Region	30.0	0.0000
Crime & Region	8.0	0.0047
Male life expectancy & Region	32.0	0.0000
Female life expectancy & Region	30.0	0.0000

Table 8:   The overall mean customer satisfaction score for the UK's Electricity Distribution Network Operators April-June 2012 (Ofgem 2013) and the effects of the adjustments suggested in Table 4.

	Satisfaction score	SDs from mean	Adjusted SDs from mean
South Wales	4.25	1.55	1.62
South West	4.25	1.55	1.62
East Midlands	4.16	1.13	1.24
Scotland - rural	4.06	0.72	0.50
West Midlands	4.05	0.68	0.65
Yorkshire	3.89	-0.04	-0.03
Southern	3.88	-0.08	-0.02
North East	3.84	-0.23	-0.24
Eastern	3.81	-0.36	-0.36
Merseyside & North Wales	3.81	-0.39	-0.44
South Eastern	3.78	-0.52	-0.49
North West	3.71	-0.82	-0.89
Scotland - urban	3.66	-1.04	-1.41
London	3.40	-2.14	-1.77