Enhancing Understanding Of Tourist Spending Using Unconditional Quantile Regression☆

S.Rudkin[a], A.Sharma*[,b]

[a]*SHU-UTS SILC Business School, Shanghai University, No.20 Cheng Zhong Road, Jiading, Shanghai, China, 201899*

[b]*Bradford University School of Management, Emm Lane, Bradford, United Kingdom, BD9 4JL*

Key words:

unconditional quantile regression, tourist spending

JEL Classifications: Z99 Z98 A97

☆We would like to thank the editor of *Annals of Tourism Research* and three anonymous referees for their valuable comments which have helped to significantly improve this paper. We would also like to thank Yang Chen, Mark Freeman, Stefano Grillini, Aydin Ozkan and Simon Peters for feedback on earlier drafts of this paper which have also helped to improve it. We are responsible for all remaining errors.

*Corresponding author. Telephone: +44 1274 234781.

Email addresses: `simon.rudkin@uts.edu.au` (S.Rudkin), `a.sharma12@bradford.ac.uk` (A.Sharma)

Enhancing Understanding Of Tourist Spending Using Unconditional Quantile Regression

Key words:
unconditional quantile regression, tourist spending
JEL Classifications: Z99 Z98 A97

An important policy goal for governments is to increase expenditures by inbound tourists, requiring appropriate statistical analysis to correctly identify important drivers of spending. In 2014 the UK received 34.4 million visits from overseas residents, netting £5.5 billion (ONS, 2015), underscoring the importance of accurate analysis. Using hitherto underutilised data from the United Kingdom International Passenger Survey (ONS, 2015) we show that past emphasis on promoting longer tourist stays misses key factors such as reason for travel (business or leisure) mainly due to inappropriate methodologies employed previously. Our central contribution is to demonstrate that conventional use of ordinary least squares (OLS) and standard quantile regressions (QR) Koenker and Bassett Jr (1978) can lead to incorrect inferences and suboptimal decisions in relation to expenditure promoting activities.

Using unconditional quantile regression (UQR) following Fortin et al. (2009), we construct quantiles of expenditure independent of the covariates, thus better accounting for underlying parameter distributions. Consequently, not only do key explanatory variables become clearer but they do so across the entire spending range. Our approach maintains benefits of QR over OLS, as reported by Brida and Scuderi (2013), but we also obtain better specified parameters by using UQR. Inferences based on UQR are more robust and should enable better decisions by policy-makers.

We utilise outbound survey data from the 2014 International Passenger Survey where interviewers administer a long survey including amounts spent, length of stay, purpose of visit, factors influencing places visited, as well as demographic details and nationality. Removing missing values yields 39,925 unique observations. Thrane (2014) reviews the major variables included within the tourism expenditure literature which informs variable choice. Owing to absence of data on income levels we employ the best available proxy. Region of origin proxies income levels here, particularly given that many countries report too few respondents to be considered individually. To avoid inferences based on insufficient observations broader categories are employed such as region of origin rather than individual countries. We also include gender and visa requirements for entry to the UK. Travelling alone, visiting friends and relatives, and being from the European Union are used as the reference categories for group size, travel purpose and region respectively, following convention. All dummies indicate the presence of a characteristic. Total expenditure and duration of stay are in logarithmic form moderating the effect of extreme high values, which is consistent with the approach suggested by Thrane (2014).

We begin by using conventional OLS techniques to assess causal relationships.[1] However, linear regressions fail to account for factors that have unequal impacts across the range of expenditure. Different variables are important for getting lower spenders to stretch their bud-

[1]Given the use of Tobits in some papers in the literature, we also estimate Tobit models. Our Tobit results are strongly in line with OLS results and are therefore not reported in the paper (available on request from the authors).

gets further as compared to enticing higher spenders to utilise more of their income (marginal utility of income differs according to levels of income). QR has been widely used in the literature to address this (Chen and Chang, 2012; Hung et al., 2012; Marrocu et al., 2015), but it has an important limitation: its parameter distributions are conditional on the choice of covariates meaning the significance, size and direction of impact can change according to specification. UQR addresses this important shortcoming by removing this dependence while effectively retaining the methodological advantages arising from QR.[2] Unconditional quantile regression permits the analysis of an outcome variable over its entire distribution instead of simply considering the mean, as is the case in OLS. OLS also fails to fully account for shifts in the underlying parameter distributions and as such incorrectly state many of the impacts that result, a problem addressed by using UQR. Though constrained by data availability, we concentrate on including the most appropriate variables, as evident from theory and the literature, while formulating our empirical models (see for example review studies by Brida and Scuderi (2013) and Thrane (2014)).

We estimate all regressions using Stata, with log spending as the dependent variable at intervals of 5% from the 10th expenditure percentile through to the 90th, but for brevity only the 10th, 50th and 90th are reported in Table 1. Significance is calculated based upon bootstrapped standard errors from 1000 repetitions, which ensures a high degree of rigour in our empirical estimations and is well in excess of the 200 repetitions used in Fortin el al (2009). In the final column we report a test of joint parameter equality between the three estimated UQR quantiles confirming significant differentials for most explanatory factors, reinforcing the inappropriate nature of OLS.

[2] We illustrate the technical details in an appendix to this paper, not meant for publication but available to referees and could be included as an online supplement.

Table 1: Regression coefficient comparison

		Quantiles of Expenditure Distribution						UQR joint coefficient equality test
		10^{th} *Percentile*		50^{th} *Percentile*		90^{th} *Percentile*		
	OLS	QR	UQR	QR	UQR	QR	UQR	
Length of stay	0.5258***	0.5366***	0.6822***	0.5593***	0.4972***	0.5615***	0.6763***	310.18***
Male	0.0941***	0.0619**	0.0835**	0.0817***	0.1138***	0.0997***	0.1197***	0.68
Aged 0-24	0.0048	0.2268***	0.1521	-0.1121***	-0.1720***	-0.1312***	0.0442	35.68***
Aged 25-64	0.2773***	0.5599***	0.7105***	0.1618***	0.1869***	0.1174***	0.1744***	50.78***
Group size: 2	-0.2337***	-0.0845**	-0.2880***	-0.2249***	-0.2368***	-0.2966***	-0.3763***	29.76***
Group size: 3 or more	-0.4173***	-0.3091***	-0.4913***	-0.4228***	-0.4742***	-0.4744***	-0.5416***	4.91
Require visa	-0.0453**	-0.6708***	-0.8642***	0.0689***	0.0947***	0.2105***	0.3181***	351.47***
Purpose: Holiday	0.4288***	0.7401***	1.2264***	0.4125***	0.4100***	0.1431***	0.0670***	510.66***
Purpose: Business	0.3267***	-0.1383*	0.1795**	0.5397***	0.4348***	0.2764***	0.3897***	20.02***
Factors influencing places visited:								
Friends, relatives or colleagues	-0.1719***	-0.1008***	0.0296	-0.2146***	-0.1980***	-0.1713***	-0.2799***	43.67***
Guidebook	0.1234***	0.1439***	0.3059***	0.0611***	0.1417***	0.0812***	0.0763*	7.29*
Review website	0.1645***	0.2244***	0.2674***	0.1392***	0.1655***	0.1327***	0.1503***	1.78
Tourist board	0.1511***	0.2026***	0.1104	0.1051***	0.2208***	0.0641*	0.0342	9.18**
Media	0.0492	0.0987	-0.0146	0.0550*	0.0478	0.0735*	0.1409	1.59
Social media	0.1707***	0.2019***	0.3787***	0.1459***	0.1147***	0.1334***	0.2808***	12.55***
Nationality region:								
North America	0.3762***	0.1167***	-0.0474	0.3807***	0.4669***	0.4973***	0.5913***	93.67***
Central America and Carribean	0.1258	0.1237	0.0248	-0.0748	0.1363	0.3061**	0.2556	0.49
South America	0.2442***	0.0554	-0.0466	0.2172***	0.2735***	0.3346***	0.5631***	18.63***
Europe: Non-EU	-0.0957***	-0.4743***	-0.7039***	-0.0335*	0.0615***	0.1574***	-0.0097	176.14***
Middle East	1.1257***	1.0233***	0.8883***	0.9828***	0.8119***	1.1991***	2.1909***	400.29***
Africa	0.6354***	0.9383***	1.0258***	0.4497***	0.5643***	0.6272***	0.7709***	20.87***
Indian Sub-continent	0.0568	-0.0241	0.0011	0.0154	0.0775*	0.1499***	0.0672	0.38
Asia	0.6281***	0.4325***	0.4389***	0.5483***	0.6168***	0.7006***	0.9980***	18.63***
Australasia	0.5000***	0.3374***	0.0549	0.4421***	0.5547***	0.5664***	0.9189***	78.49***
Constant	4.6383***	3.0902***	2.5242***	4.8081***	4.8264***	5.8879***	5.9839***	
N	39525	39525	39525	39525	39525	39525	39525	

UQR coefficient equality test is a joint test that coefficients from the three UQR models are identical. Significance denoted by $^{*}(p < 0.05)$,$^{**}(p < 0.01)$,$^{***}(p < 0.001\}$

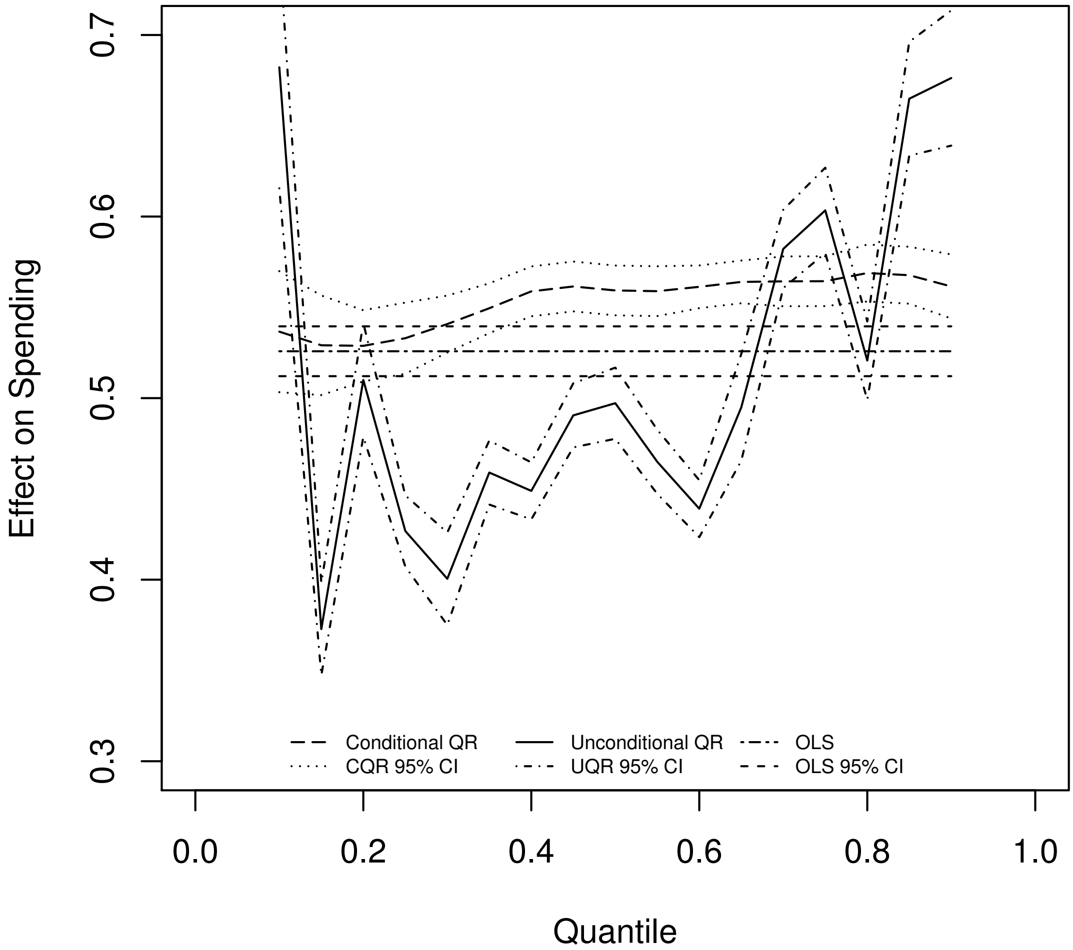

Figure 1: Length of stay

Table 1 shows that many coefficients are significant. More crucially it indicates a differential between explanatory factors such as the influence of the tourist board and guidebooks, which are significant at the lower end, and visitors being from North America which impacts only higher up the spending distribution. However it is in the comparison between QR and UQR that the advantage of UQR becomes evident. In most cases the range of coefficients is much greater using UQR and this better focuses attention on the quantiles where that particular variable is most influential. To illustrate our argument for UQR 'Purpose: Holiday' has a QR coefficient of 0.7401 at the 10th expenditure percentile and 0.1431 at the 90th, while the corresponding UQR values are 1.2264 and 0.0670 respectively, highlighting greater coefficient variations using UQR which should enable target policy to be better targeted on larger potential impacts. The economy has more to gain from holidaying tourists relative to those visiting family than either OLS or QR suggests. Similar conclusions can be drawn from other variables in Table 1.

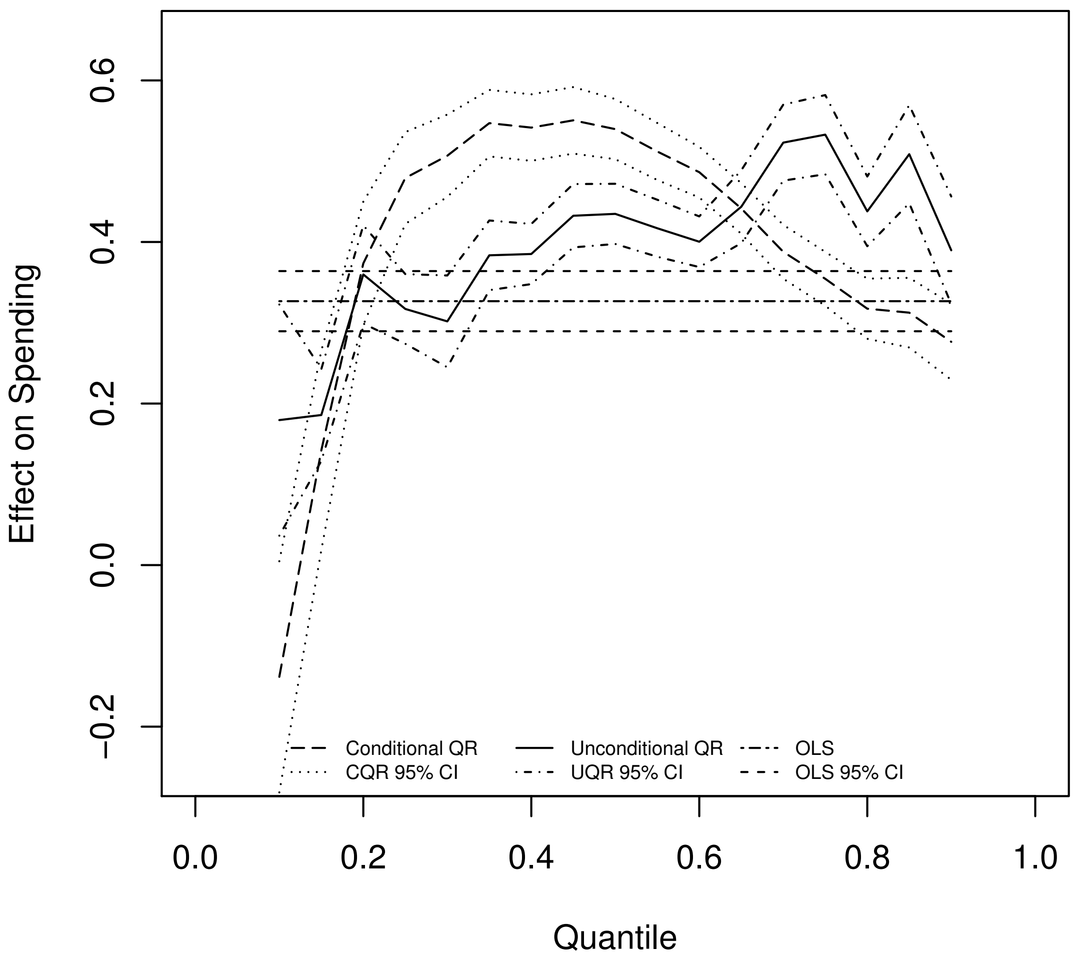

Figure 2: Business travellers

Further illustration of the benefits of UQR can be seen in Figures 1 and 2, where coefficients on the length of stay (Figure 1) and business being the main visit purpose (Figure 2) are plotted with 95% confidence intervals. These variables are selected as they receive extensive coverage in the expenditure literature (Thrane, 2014). QR is assumed to be able to untangle the averaging effect of OLS, but here QR reports values consistently above OLS; UQR does display averaging however. Both QR and OLS overstate the role of length of stay. At the 10th percentile QR suggests a negative coefficient on 'Purpose: Business' while UQR and OLS are positive, indicating incorrect conclusions that QR can bring about due to conditioning especially within the distribution tails. In contrast, business travellers' spending is consistently underestimated by OLS, but QR overstates these at lower quantiles and understates these after the 65th percentile. It is intuitive to promote longer stays to achieve higher expenditures overall and entice more business travellers but UQR allows policy-makers to see to what extent this will be valuable, particularly the implication that the focus should be more on attracting corporate visitors and less on the duration of visits.

This note highlights the value of using UQR for addressing the limitations inherent within previous methods involving conditional parameter distributions for spending analysis (QR and OLS). Using unique data and robust analysis using improved methods, our paper clearly demonstrates the over-importance attached to length of stay and the inadequate attention given to business travelers in previous research. There are clear benefits from UQR's methodological robustness for assessing the multitude of variables related to tourist expenditures, particularly given UQR's ability to inform across the spending distribution. Given tourism's importance for the UK it is critical for expensive promotional activities to be targeted efficiently for ensuring effective policy making.

References

Brida, J. G. and Scuderi, R. (2013). Determinants of tourist expenditure: a review of microeconometric models. *Tourism Management Perspectives*, 6:28–40.

Chen, C.-M. and Chang, K.-L. (2012). The influence of travel agents on travel expenditures. *Annals of Tourism Research*, 39(2):1258–1263.

Fortin, N., Lemieux, T., and Firpo, S. (2009). Unconditional quantile regression. *Econometrica*, 77(3):953–973.

Hung, W.-T., Shang, J.-K., and Wang, F.-C. (2012). Another look at the determinants of tourism expenditure. *Annals of Tourism Research*, 39(1):495–498.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Marrocu, E., Paci, R., and Zara, A. (2015). Micro-economic determinants of tourist expenditure: A quantile regression approach. *Tourism Management*, 50:13–30.

ONS (2015). International Passenger Survey [data collection] SN: 7534.

Thrane, C. (2014). Modelling micro-level tourism expenditure: recommendations on the choice of independent variables, functional form and estimation technique. *Tourism Economics*, 20(1):51–60.

Technical Appendix - To be included as an online supplementary note only

Quantile regression (see Koenker and Bassett Jr (1978)) permits the analysis of an outcome variable over its entire distribution instead of simply considering the mean, as is the case in Ordinary Least Squares regression (OLS). OLS regressions also fail to fully account for shifts in the underlying parameter distributions and as such incorrectly state many of the impacts that result. The relevance of quantile regressions is best captured by the following quote (Mosteller and Tukey, 1977: 266):

> What the regression curve does is a grand summary for the averages of the distributions corresponding to the set of xs. We could go further and compute several different regression curves corresponding to the various percentage points of the distribution and thus get a more complete picture.

We denote the outcome variable as Y, with y being the value corresponding to one particular observation within the sample. For each data point there will also be a corresponding set of explanatory variables X. For the observation with output variable value y the corresponding explanatory variable values are collected as x. As there may be more than one explanatory variable for each observation, X is a matrix of observed values and x is a vector. Koenker and Bassett Jr (1978) posit that coefficients for the explanatory variables could be estimated at quantile τ, $\tau \in (0,1)$ by solving:

$$\underset{\beta \in \Re^K}{\text{minimize}} \quad \left[\sum_{t \in (t: y \geq x_t \beta)} \tau \left| y_t - x_t \beta_\tau \right| + \sum_{t \in (t: y \leq x_t \beta)} (1-\tau) \left| y_t - x_t \beta_\tau \right| \right] \tag{1}$$

Here β_τ is the vector of coefficients that we wish to estimate. The subscript t is added because the first term considers only those observations where the outcome variable is above the stated quantile, whilst the second term looks at only those for which y is below the given quantile. Observations are penalised according to their distance from the desired τ but all observation are included when estimating every set of parameters. Critically equation (1) is dependent on the conditional variables, and hence the Koenker and Bassett Jr (1978) technique has come to be known as Conditional Quantile Regression (CQR).

In our analysis, we make an important contribution to the literature by implementing Unconditional Quantile Regression (UQR) for the first time for the study of tourist expenditure. This is a variable of significant interest within tourism research but in our view previous empirical methods that have been used (OLS and QR in particular) suffer from important limitations which are addressed by using UQR. The method follows Fortin et al. (2009) and it begins by formulating a Recentered Influence Function, RIF, which is constructed without reference to the covariates. Fortin et al. (2009) propose that:

$$RIF\,(Y;q_\tau,F_Y) = q_\tau + \frac{\tau - \mathbb{1}\{Y \le q_\tau\}}{f_Y\,(q_\tau)} \tag{2}$$

In this result the first of the additional important terms is q_τ, which is the observed τ quantile, for example the 75th percentile of tourist expenditure. F_Y is the distribution of the variable Y, whilst f_Y is the marginal distribution at q_τ. In our analysis these are the distribution and marginal distribution of observed expenditure. $\mathbb{1}\{Y \le q_\tau\}$ is an indicator function which takes the value one when the observed value of Y for a particular observation is below the τ^{th} quantile.

Using these RIF values enables us to perform a regression on the covariates that may be estimated using standard OLS techniques. Despite applying the simple regression the construction of the RIF ensures that the coefficients that are provided from the regression are unique to the particular quantile of the Y distribution that was used to calculate the RIF. In estimating this regression of the RIF on the explanatory variables we apply bootstrapped standard errors to provide a greater robustness than is the case for standard linear modelling applied widely within the literature. The significance of explanatory variables in explaining the observed outcomes at different levels on the Y distribution is thus calculated without the choice of variable being integral in the calculation of that output variable distribution.

For the interested applied researcher both CQR and UQR can be estimated readily by most software packages, with Stata being used in our analysis, making use of adapted code from Fortin et al. (2009). The advantages of CQR over OLS are well established, but in our analysis we demonstrate that UQR is an important improvement in terms of empirical modelling and the reliability of estimates so obtained, which in our view must be considered by researchers and practitioners alike.

Reference:

Mosteller, F. and Tukey, J.W. (1977). *Data analysis and regression: A second course in statistics.*, Addison Wesley, Reading: Mass edition.