NORTHERN ARIZONA UNIVERSITY
The W. A. Franke College of Business

Quantile Regression Analysis of Visitor Spending:
An Example of Mainland Chinese Tourists in Hong Kong

Working Paper Series—09-06 | April 2009

Pin T. Ng
Northern Arizona University
The W. A. Franke College of Business
Box 15066
Flagstaff, AZ 86011-5066
pin.ng@nau.edu

and

Alan A. Lew
Northern Arizona University
Department of Geography, Planning and Recreation
Box 15016
Flagstaff, AZ 86011-5016
alan.lew@nau.edu

Quantile Regression Analysis of Visitor Spending:
An Example of Mainland Chinese Tourists in Hong Kong

Tourism is a *basic export industry* (Hall and Lew 2009). It is a basic industry because almost all of the income generated by tourism comes from local products and experiences sold to customers who pay for them with money earned outside of the local community. Traditional basic industries include agriculture, mining and manufacturing to the degree that they are exported and sold to people and entities who earn their livelihood outside of the community. Tourism is a form of export industry, even though its products are consumed *in situ* (at the site of production) and are not physically exported. What is important is that a purchase is made of local products (or experiences) by consumers who normally reside outside of the community. Basic export industries are essential to the economic well-being of a community, because a place would wither and die without some source of external income.

The argument that tourism activities effectively increase the gross wealth or income of a community is what makes tourism a popular form of economic development (Frechtling 1994a). Economic leakages, such as when tourists purchase products that are imported into the local economy, can significantly decrease the potential benefit of tourism industry expenditures. In fact, much of what a tourist purchases has a high import content, from transportation and fuel on the trip from their home, to accommodations made from imported steel and wood, to restaurant foods and souvenirs grown and made in distant lands.

Despite these challenges, the essence of the tourism economy is built on the expenditures made directly by tourist in the destination. Segmenting the tourist market based on expenditure patterns is a common approach to understanding the economic impacts of tourism on a destination (Weber 1995). Such segmentation is often undertaken with the goals of assessing visitor destination or shopping satisfaction (Heung and Cheng 2003; Wong and Law 2003; Joppe, et al. 2001), identifying "big-spenders" (Legohérel and Wong 2006; Díaz-Pérez, et al. 2005; Wicks and Schuett 1993), or trying to determine if one part of a tourist population spends more or differently than another part (Agarwal and Yochum 1999). For many such studies, market segmentations are simply based on country of origin, and the variations in expenditures are simply the overall mean or purchase category means for the respondent groups (Suh and Gartner 2004).

For example, Rosenbaum and Spears (2005) used analysis of variance (ANOVA) to examine variations in expenditure patterns for among different nationalities visiting Hawaii and found that Japanese visitors were not only the biggest spenders, by far, but also planned to stay longer in Hawaii and spend more time shopping there than any other group during the survey period. They also found that repeat visitors planned to spend more in Hawaii than first time visitors. Other common independent variables used to segment market populations include income, age and gender (Lehto, et al 2004; Mok and Iverson 2000), tourist interests (Mehmetoglu 2007) and trip type (Oh, et al. 2004). A separate set of studies consists of those that model changes in the overall tourist expenditures at a destination over time (Narayn 2005; English 2000).

In most of the tourist expenditure studies reviewed above, the statistical significance of differences between market segments was based on least-squares regression. However, the least-square regression is designed to estimate only the *average* (mean) spending behavior across groups of tourists. Since not all tourists behave like an *average* tourist, it is valuable to find out not only the *average* difference across groups but also the differences between the higher and lower ranges across the groups. For our analysis of tourist expenditures, this is the difference between the *big* spenders and the more *frugal* spenders after controlling for variations in demographic characteristics. Knowing the behavior of big spenders in comparison to frugal spenders can help tourism planners achieve the highest impact in their advertising expenditures and travel programs.

The standard deviation, range or other measures of dispersion of the distribution of a spending category can provide a rough indication of the degree of spread of extreme values around the mean without taking into account the effect of other factors like the demographic characteristics. A least-squares regression can control for the effect of these additional independent variables on the *average* spending category. However, a richer and more precise understanding can only be achieved through quantile regression analysis, which allows examining and comparing different levels of response in a spending category given the variation in the independent variables. Respondents who only spent, for example, 10% of the mean on meals outside of their hotel, can be directly compared with those who spent twice the mean, and those who spent exactly the mean amount holding all the other factors fixed.

Using quantile regression, we were able to see the different spending behavior across the whole spectrum of the tourist population, from the big spenders, through the moderate spenders, and to the frugal spenders among Mainland Chinese visitors to Hong Kong. We have found that the number of previous visits had significant increasingly positive effects on the middle 50% of the shopping spending distribution. Being a repeat visitor affected shopping spending more for heavier spenders than for the more frugal. It also had increasing positive impacts on local transportation spending over the whole distribution. These findings support and expand the results that other studies in other destinations have found. The average Mainland Chinese repeat visitor to Hong Kong was found to spend more overall during their trip than did first time Mainland Chinese visitors, especially on transportation.

Duration of stay, however, was found to not have a significant effect on shopping spending behavior overall. It was also found to not have a significant impact on spending for meals outside hotel and hotel bills for an average spender although it did have significantly increasing positive effects on the upper 60% of the distribution in spending for meals outside hotel and spending for local transportation, and increasing positive influences on the top 30% of the hotel spending distribution. Thus, efforts to increase the stay of Mainland Chinese visitors would mostly affect the amount of money they spend on food outside of their hotel and on transportation, at least among the middle and higher spenders. The high-end hotels would also benefit more than average from longer stays by their Mainland Chinese clientele.

Visitors who were proprietors/owners also spent more on shopping, meals outside hotels and hotel bills than visitors in other occupation categories for most of the distributions. Shopping centers, restaurants and hotels in Hong Kong could more effectively increase their revenue by targeting this particular occupational group of visitors. The “age 26 to 40” group of visitors had higher spending on shopping than the other age groups across the whole distribution, and the top 10% of the spenders (i.e., the “big spenders”) among them spent ten times as much as the bottom 10% (“frugal spenders”), when compared to the other age groups. This suggests that focusing the shopping promotional campaign on this particular group of Mainland Chinese visitors could have a greater impact on Hong Kong tourist expenditures.

The rest of the paper is organized into the data and methodology section, followed by the section on the major findings, and the last section of conclusion and discussions on implications for policy makers and entities.

Data and Methodology

The data used in this study were the same as those used in Wang (2004), which were extracted from the visitors’ survey conducted by the Hong Kong Tourist Board (HKTB) in 1999. The 634 Mainland Chinese vacationers aged 16 or above interviewed face-to-face were randomly selected at four border control points: Hong Kong International Airport, Hung Hom Railway Station, China Hong Kong Ferry Terminal and the Macao Ferry Terminal. The number of respondents interviewed at each of the four control points was proportionate to the actual number of visitors that passed through the control points. The survey was implemented on a continuous basis, with a relatively equal number of interviews conducted in each month of the year.

Data on marital status were classified into “married”, “never married” and “other”, which was used as the base so that the estimated regression coefficient for marital status should be interpreted relative to this category. The “length of stay” measured the number of nights the visitors stayed in Hong Kong during their current trip. Repeat visitors were identified by those whose “number of visits” to Hong

Kong was at least two. Education level was classified into three levels: “primary or no education”, “secondary/high school” and the base level “college or above”. Occupations were characterized by four dummy variables: “professional/managerial”, “proprietor/owner”, “junior white collar”, and “blue collar” with the base level capturing “other job types”. Age group was broken down into “age 16-25”, “age 26-40”, “age 41-60” and “greater than 60”, which was used as the base. Gender was split into “female” (as the base) and “male”.

To study tourist behavior on four major spending components (shopping, meals outside of hotels, local transportation and hotel, all measured in US dollars), Wang (2004) performed separate least-squares regressions of the four components on the common set of independent variables that captured visitors' socio-economic characteristics (marital status, age, gender, occupation and educational attainment), the length of stay and the number of previous visits. His least-squares regression results were presented in Table 7 of Wang (2004). Investigation of the variance inflationary factor (VIF) of the independent variables, however, revealed that the “never married” dummy variable had the highest VIF value of 17.29, which suggests high correlation between this dummy variable and other independent variable(s). After dropping this dummy variable, we found that the dummy for “age 26 to 40” had the highest VIF value of 12.28 among the remaining VIF values followed by the VIF value of 10.46 for the “age 41 to 60” dummy. We decided to drop the dummy for “age 41 to 60” so that the base level for the age dummies became “age 41 and above”. After this, none of the remaining independent variables had a VIF value higher than 2.0 and there was no more evidence of multicollinearity among the remaining independent variables.

We performed a special case of the White test for heteroskedasticity by regressing the squared-residuals on the fitted dependent variable and its squared term. The chi-square statistic when using the “total spending”, “spending on shopping”, “spending on meals outside of hotels”, “spending on local transportation” and “spending on hotel” in turn as the dependent variable is, respectively, 19.20, 19.39, 33.34, 58.40, and 6.81. They all have essentially zero *p*-value. Thus, there is extremely strong evidence of heteroskedasticity (a predictable change in variance over the distribution of the dependent variable) in the least-squares regression models. To make sure that the heteroskedasticity test does not pick up the effect of functional form misspecification in the regression model, we performed the test by including the quadratic terms of “length of stay” and “number of visits” and logarithmic transformation on the dependent variable. The conclusions remained the same. Hence, we computed the Eicker-Huber-White heteroskedasticity-robust standard errors in all the least-squares regression results below, which should be more reliable than those reported in Wang's (2004) Table 7.

Because of the presence of heteroskedasticity, quantile regression becomes even more relevant for revealing any potential variation in the effect of the independent variables on the dependent variables over the various segments of the population. If there is no variation in the dispersion of the dependent variable across the different ranges of the covariates, the conditional mean is capable of revealing information on other quantiles of the conditional distribution with the addition of some parametric assumptions. In the presence of heteroskedasticity of any form, however, no parametric distribution model can reasonably be expected to capture the unknown conditional distribution to be useful for the estimation of the covariate effects in the other quantiles using the least-squares regression. Quantile regression, on the other hand, is designed to capture any potential covariate effects in any chosen quantile and does not rely on any parametric specification of the conditional distribution.

Given *n* observations of the dependent variable y_i, and *k* independent variables represented by the *k*-vector x_i for *i* = 1, …,*n*, the *k*-vector of τ –th quantile regression coefficients, $\beta(\tau)$, minimizes

$$\min_{\beta(\tau)} \sum_{i:y_i - x_i\beta(\tau)\geq 0} \tau\left|y_i - x_i\beta(\tau)\right| + \sum_{i:y_i - x_i\beta(\tau)<0} (1-\tau)\left|y_i - x_i\beta(\tau)\right|$$

where $0 < \tau < 1$ determines the desired conditional quantile of interest. In the objective function above, all the positive residuals receive a weight of τ while the negative ones receive a weight of $(1-\tau)$. Hence, $100\tau\%$ of the dependent observations will fall above the τ-th quantile regression hyperplane and $100(1-\tau)\%$ below. For $\tau = 0.5$, the quantile regression hyperplane bisects the dependent variable into two halves such that half of the observations fall above while the other half below the regression hyperplane and yield the *median* regression estimates as a special case. Hence, any one of the *k* components of the quantile regression coefficients, $\beta_j(\tau)$, provides an estimate of the marginal effect of the associated independent variable x_j on the dependent variable for the τ-th quantile of the cohort holding the effects of the remaining independent variables fixed.

All the quantile regression coefficients in this study were computed using the **quantreg** package (Koenker, 2009) available for the GNU Free Software **R** for statistical computing and graphics (R Development Core Team, 2008). The algorithm uses a modified version of Barrodale and Roberts's (1974) algorithm for L_1 regression as described in Koenker and d'Orey (1987, 1994). The standard error presumes local linearity of the conditional quantile functions and computes a Eicker-Huber-White sandwich estimate using a local estimate of the sparsity. See Koenker and Hallock (2001) for an excellent non-technical introduction to quantile regression.

Results

The least squares regression (LSR) and quantile regression (QR) estimated coefficients are presented in Tables 1 to 5 for the dependent variables: "spending for shopping", "spending for meals outside hotel", "spending for local transportation", "total hotel bill" and "total spending", respectively. Since there is a plethora of numbers in the tables, we present the information graphically in a more parsimonious way in Figures 1 to 5. The tables show the 9 distinct deciles representing 10% increments in τ along the distribution. Each panel in a figure presents the quantile regression estimates for the coefficients of a covariate. For each of the covariates, 17 distinct quantile regression estimates for τ ranging from 0.1 to 0.9 in increments of 0.05 are plotted as solid dots joined by a solid line. The vertical distance of a solid dot from the horizontal axis can be interpreted as the impact of a one-unit change in that independent variable on the dependent variable for that chosen τ –th quantile of the dependent variable holding other independent variables fixed. For example, in Figure 1, the second panel in the first row shows the relationship between the independent variable "number of visits" and the dependent variable "spending on shopping". There is an increasing impact on "spending on shopping" with each additional number of visits for the lower 0.70 quantiles (reflecting the lower 70% of shopping spending) and a decreasing impact beyond that. Similarly, the second panel in the second row shows that visitors who are "proprietors" spend more than other occupation types among the top half (0.50) of the population in terms of spending on shopping. The top 10% of "proprietor" shoppers spent about $9,000 more than "other job types" visitors, which is the base level of the occupation dummy variables.

The two solid curves that envelop the solid dots are the 95% confidence band for the quantile regression coefficients. Hence, a confidence band corresponding to a solid dot of a specific τ that does not contain the solid horizontal x-axis represents a quantile regression coefficient that is statistically significant at the 5% level for that particular τ. In the example of the second panel in the top row in Figure 1, the "number of visits" has a significant impact on shopping bills statistically only for the middle 40% of the population strata ($0.3 < \tau < 0.7$) in which the confidence band does not envelop the solid horizontal x-axis. For the rest of the population strata, the "number of visits" should be considered as not having an impact on shopping spending, even though the estimated coefficients are positive, since they are not significant statistically at a 5% level.

The dash horizontal line in each panel represents the least-squares estimate of the conditional mean effect for the entire distribution, while the two dotted lines represent the 95% confidence interval constructed using the heteroskedasticity-robust standard error. Hence, the least-square estimate is statistically significant at the 5% level if the dotted lines do not envelop the horizontal axis. In the example from the second panel in Figure 1, the "number of visits" does not have a significant impact on average shopping bills because the two dotted lines envelop the horizontal axis.

In the following sections, we highlight the more significant differences that can be detected in applying the quantile regression method to data that was originally assessed by Wang (2004) using a more traditional last-squares estimate approach. This will be discussed through the results shown in Tables 1 to 5, and correspondingly Figures 1 to 5, which cover spending on shopping, meals outside a hotel, local transportation, hotels, and total spending, respectively.

Spending for Shopping

Wang (2004) found that the "number of visits" had a statistically significant positive effect on "spending for shopping" for the *average* Mainland Chinese spender in Hong Kong according to his least-squares estimate. However, our least-squares estimate in Table 1 concludes otherwise. This is due to the larger heteroskedasticity-robust standard error used in computing our *t* test statistic. The *t* test statistic was indeed significant at the 5% level if the traditional standard error was used. The quantile regression estimates reveal additional insight. This is shown graphically in the second panel in the top row of Figure 1, where the "number of visits" has an increasingly marginal effect on shopping expenditure as we move from the first quartile of the distribution where $\tau = 0.25$ towards the third quartile when τ reaches 0.75 but not for the bottom and top 25% shoppers (Table 1 corresponds directly to Figure 1). Therefore, each additional visit only has a higher impact on shopping expenditure among the middle 50% shoppers. While this reaffirms the previous findings that higher number of previous visits results in higher shopping spending, it is true only for the middle 50% of shoppers. The *average* effect, which is unduly influenced by the outliers in the upper and lower tails of the distribution, turns out to be statistically insignificant.

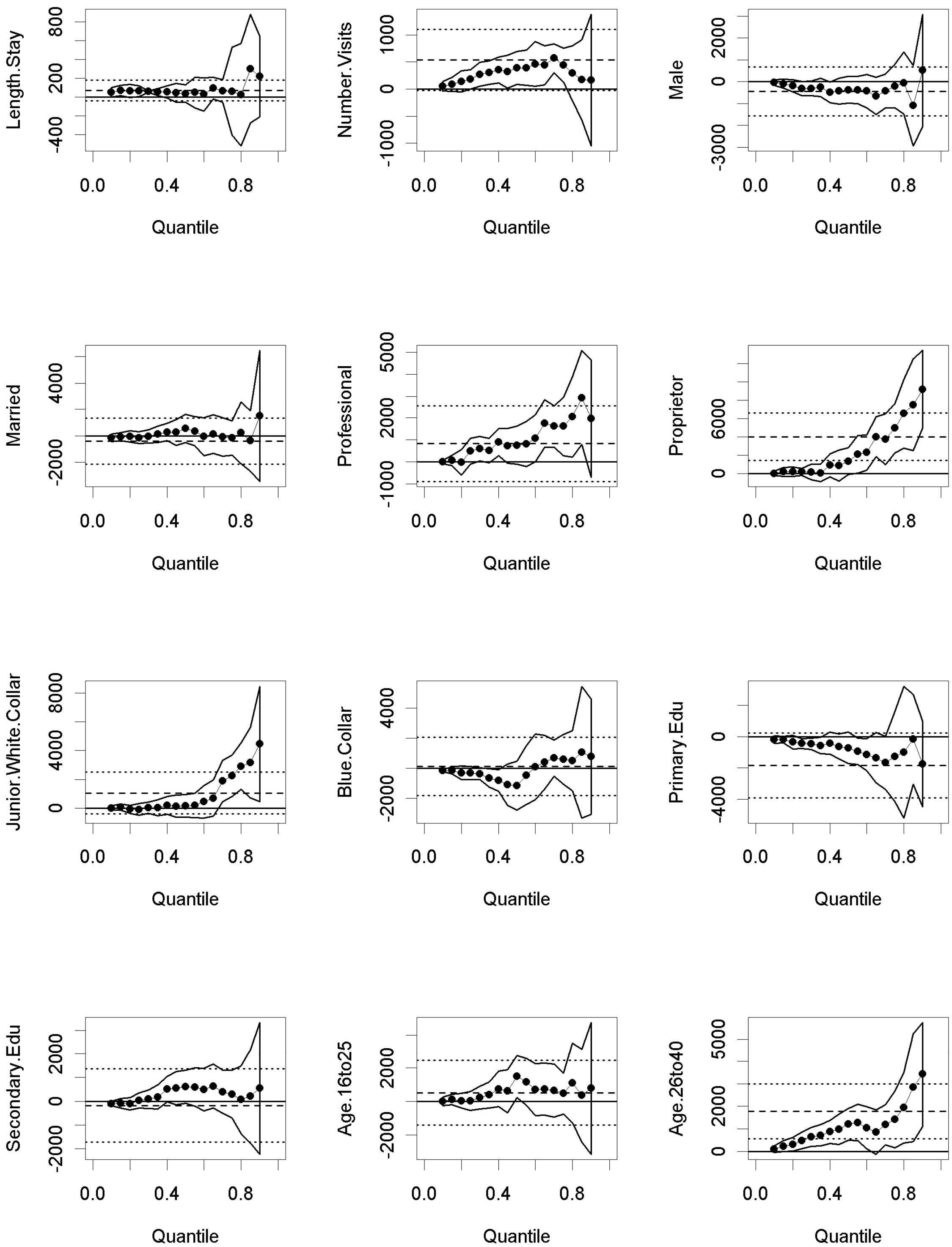
Wang's original least-squares estimate showed that an *average* "proprietor/owner" spends about US$4,000 more than people in other job types, holding all other factors constant. An investigation of the quantile regression estimates (Table 1 and Figure 1, second row, second panel), however, reveals that "proprietors/owners" spend more than other job types only for the upper half of the shopping spending distribution, with the discrepancy in spending increasing as we move towards the upper tail of the distribution. The upper 10% among the proprietors/owners spend slightly more than US$9,000, which is more than double that of an *average* spender, when compared to the other job types.

In our final example, the *average* "age 26 to 40" group was shown by Wang to spend slightly under US$2,000 more than the other age group according to the least-square estimate. The quantile regression estimates (Figure 1, bottom row, third panel) also reveal that the higher spending effect of this group increases as τ increases throughout the whole spectrum of the distribution, with a couple of less than significant exceptions. The higher shopping expenditure of this particular age group might have reflected the fulfillment of a "conspicuous consumption" urge (spending for status) that is met by the earning ability of the group, as observed by Charles, et al. (2007).

Table 1. Estimated coefficients for the least-squares regression (LSR) and quantile regressions (QR) for "Spending on Shopping". The number of asterisks signifies the level at which the coefficients are significant with 0.05 (*), 0.01 (**) and 0.001 (***).

Variables	**LSR**	**QR**			
		.1	**.2**	**.3**	**.4**
Intercept	2172.12	135.00	237.50	297.27	110.44
Length of Stay	69.93	48.33*	64.40	57.27***	52.31
Number of Visits	535.25	55.00	140.70	268.73*	350.44**
Male	-442.70	-31.67	-178.80	-305.27	-485.05*
Married	-391.11	-155.00	-51.40	-48.00	282.75
Professional	836.23	0.00	-14.00	604.00*	904.62
Proprietor	3999.21***	31.67	221.60	178.00	920.00
Junior White Collar	1055.20	10.00	-78.40	21.81	184.95
Blue Collar	116.22	-161.67***	-297.20	-359.46	-799.56*
Primary /No Education	-1829.08	-193.33	-341.20	-462.54*	-415.49
Secondary/High School	-177.71	-100.00*	-108.40	89.46	516.81
Age 16 to 25	519.39	-6.67	42.60	233.12	750.00
Age 26 to 40	1775.95*	100.00	310.00*	637.46**	864.51***
	QR				
	.5	**.6**	**.7**	**.8**	**.9**
Intercept	238.16	1505.00	1927.50*	2498.89	2838.61
Length of Stay	36.12	26.86	62.50	22.78	217.22
Number of Visits	389.59*	461.51*	567.50***	291.11	162.91
Male	-369.39	-431.05	-420.00	-54.44	514.05
Married	550.00	-58.49	-82.50	227.22	1496.84
Professional	764.49	1076.74*	1617.50***	2050.56*	1976.33
Proprietor	1358.98	2290.23*	3732.50**	6527.22***	9184.30***
Junior White Collar	155.51	450.47	1882.50*	2913.33***	4462.41*
Blue Collar	-1147.14	85.00	675.00	500.00	767.09
Primary /No Education	-713.88	-1147.09*	-1667.50	-985.56	-1746.84
Secondary/High School	614.49	492.56	400.00	68.89	553.16
Age 16 to 25	1494.49*	714.65	647.50	1113.89	796.84
Age 26 to 40	1204.90***	1045.70*	1182.50**	1941.11*	3427.97**

Figure 1: Least-squares and quantile regressions estimates for “Spending on Shopping.”



Note: The horizontal x-axis shows the 17 distinct τ values (ranging from 0.1 to 0.9 in 0.05 increments) that determine the specific quantiles of interest. The vertical distance from the x-axis of each dot for the corresponding τ represents the magnitude of that specific quantile regression estimate. The solid curves that envelope the dots show the 95% confidence band for the 17 quantile regression estimates. The quantile regression estimate for a particular τ is considered statistically significant at the 5% level when the x-axis is outside of the envelope for that particular τ. The single thin dashed horizontal line is the least-squares estimate of the entire distribution. The two dotted lines are the 95% confidence band for the entire distribution and the least-squares estimate is statistically significant at the 5% level when the x-axis is outside the two dotted lines.

Spending for Meals Outside Hotel

The "length of stay" was found to be statistically significant in affecting "spending for meals outside hotel" for an *average* Mainland Chinese spender in Hong Kong by Wang (2004). Our Table 2 (and Figure 2, first row, first panel), however, shows otherwise. Again, this is due to the difference in standard error between Wang's (2004) and our study. The quantile regression estimates from Figure 2 do suggest statistically significant positive effect from "length of stay" for the spenders, but only for spenders who fall in the upper 60% of the distribution. Also the positive marginal effect increases considerably as we move towards the upper tail of the distribution, with the top 10% spending almost five times more (US$440) than the median spenders (US$88).

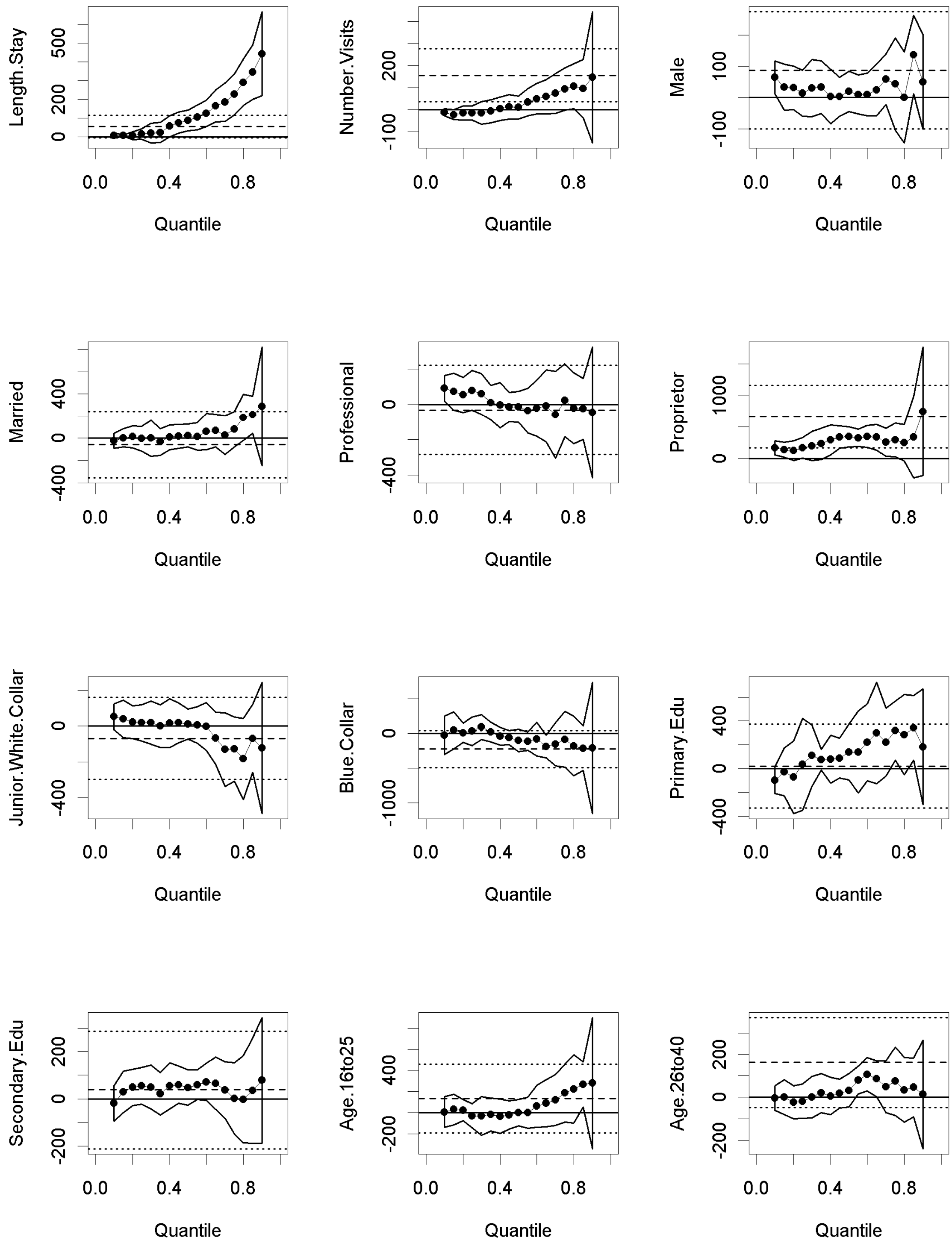
The traditional least-squares estimate indicates that the "number of visits" has a statistically significant positive impact on spending for meals for an *average* visitor. However, the quantile regression estimates (Table 2 and Figure 2, first row, second panel) show that this positive impact only exists among spenders at the 0.8 quantile of all spenders. A closer investigation of the data revealed that there was an outlying visitor who spent US$20,000 on meals in 6 days on a fifth visit to Hong Kong. This influential observation had an undue influence on the least-square estimate and caused it to become statistically significant. If this outlier were removed, even though there is a slight trend toward higher spending with increased visitations, there is actually no statistically significantly difference between first time and repeat visitors in this expenditure category. This highlights the sensitivity of the standard least-square estimates results to outliers, while showing how the quantile regression estimates are more intrinsically robust.

Wang found that average "Proprietors/owners" spend about US$720 more than other job types while our estimate indicates about US$666 more. They do spend around US$300 more than other job types for τ between 0.4 and 0.7 (Table 2 and Figure 2, second row and third panel).

Table 2. Estimated coefficients for the least-squares regression (LSR) and quantile regressions (QR) for "Spending on Meals Outside Hotel". The number of asterisks signifies the level at which the coefficients are significant with 0.05 (*), 0.01 (**) and 0.001 (***).

Variables	**LSR**	**QR**			
		.1	**.2**	**.3**	**.4**
Intercept	-	-	184.36*	-	290.00*
Length of Stay	-	-	-	-	-
Number of Visits	155.85*	-	-	-	-
Male	-	64.30*	-	-	-
Married	-	-	-	-	-
Professional	-	92.90*	-	-	-
Proprietor	666.42*	170.70**	-	-	296.09*
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	-	-	-
Age 26 to 40	-	-	-	-	-
	QR				
	.5	**.6**	**.7**	**.8**	**.9**
Intercept	258.00**	-	-	-	-
Length of Stay	88.00**	125.45***	185.26**	291.25***	443.87***
Number of Visits	-	-	-	106.25*	-
Male	-	-	-	-	-
Married	-	-	-	-	-
Professional	-	-	-	-	-
Proprietor	344.00***	349.09***	260.66*	-	-
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	-	-	-
Age 26 to 40	-	106.36**	-	-	-

Figure 2: Least-squares and quantile regressions estimates for “Spending on Meals Outside Hotels.”



Note: Each dot represents a 5% quantile of the distribution. The solid curves that envelope the dots show the 95% confidence band for each 5% quantile dot and is significant when the "0" axis is outside of the envelope. The single thin dashed horizontal line is the least-squares mean of the entire distribution. The two dotted lines are the 95% confidence band for the entire distribution and is significant when the "0" axis is outside the two dotted lines. The 5% quantiles correspond directly to the 25% quartiles in Table 2.

Spending for Local Transportation

Similar to Wang (2004), our results in Table 3 and Figure 3 (first row, first panel) also show that the longer an *average* Mainland Chinese tourist stays in Hong Kong, the more they will spend on local transportation. The quantile regression estimates reveal that this positive impact becomes larger as we move up the distribution throughout the whole spectrum of τ. For example, the top 10% of spenders on local transportation spend about five times more than the *average* spenders while the bottom 25% spend about one-fifth of that of the *average* spenders.

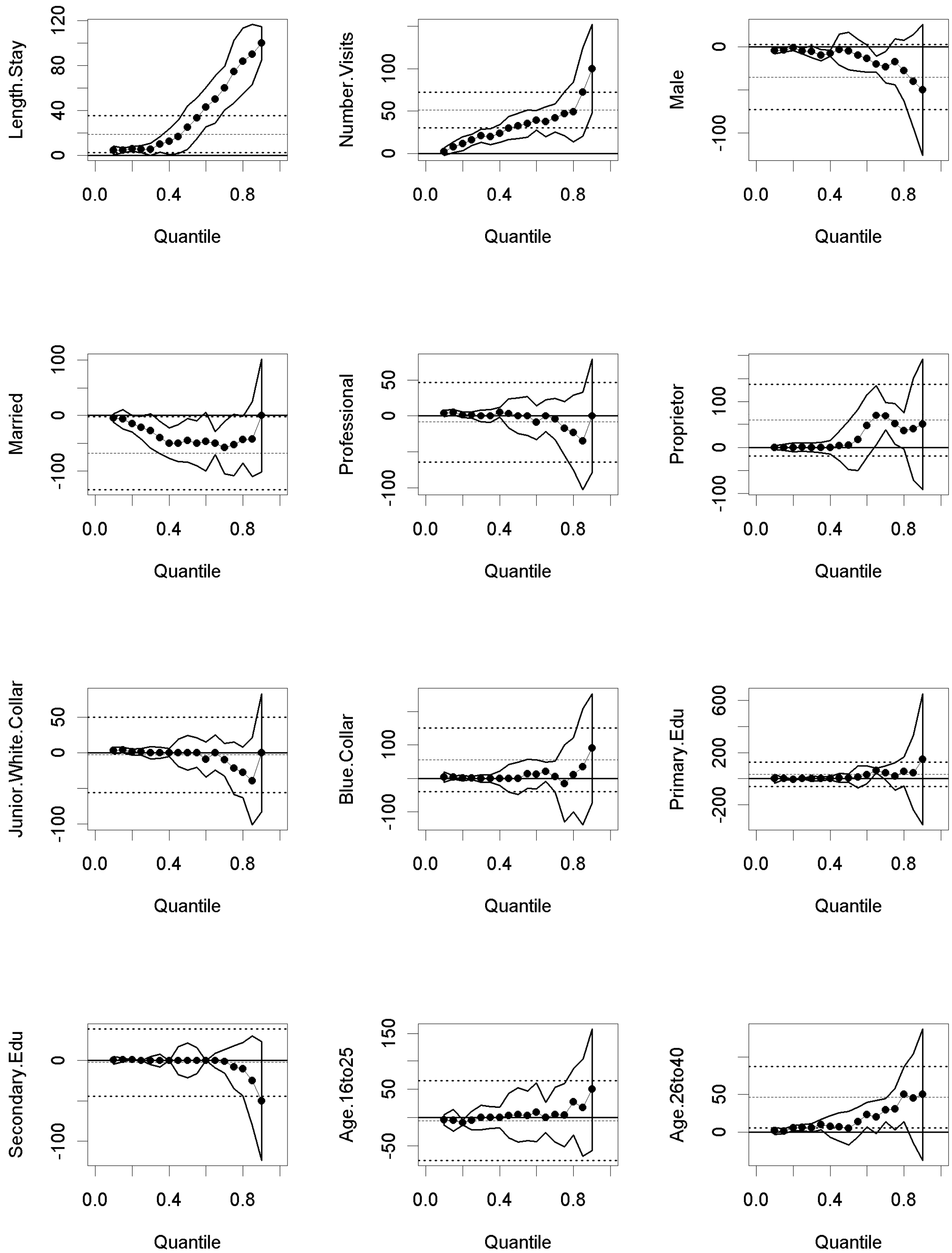
Similarly, the "number of visits" (Table 3 and Figure 3, first row, second panel) has a statistically significant positive impact on local transportation expenditure, which corresponds to Wang's findings. The positive impact also increases as τ increases. Each additional visit brings about twice as much spending on local transportation for the top 10% when compared to the *average* visitors after taking all the other factors into consideration.

As Wang found, visitors who are married (Table 3 and Figure 3, second row, first panel) spend about US$70 less than their counterpart on *average*. The quantile regression estimates show a statistically significant negative impact among the married tourists only for the 0.35 to 0.8 quantiles of the distribution. Hence, the 35th to 80th percentiles of the spenders among married tourists spend less on local transportation compared to their non-married counterparts. As noted above, the quantiles are superimposed on the married/not married variable. Similar to the impact on spending for shopping, the "age 26-40" group also has a higher *average* spending on local transportation. This positive impact is also statistically significant for the 0.6 to 0.8 quantiles. The 60th percentile spenders among this age group spend about US$23 more while the 80th percentile among this group spend US$50 more than the "age 41 and above" group.

Table 3. Estimated coefficients for the least-squares regression (LSR) and quantile regressions (QR) for "Spending on Local Transportation". The number of asterisks signifies the level at which the coefficients are significant with 0.05 (*), 0.01 (**) and 0.001 (***).

Variables	**LSR**	**QR**			
		.1	**.2**	**.3**	**.4**
Intercept	-	-	-	-	-
Length of Stay	19.10***	4.61*	6.10***	5.55*	12.50*
Number of Visits	51.35***	-	11.71**	20.83***	23.75***
Male	-	-4.23*	-	-	-7.50***
Married	-68.36*	-	-	-	-50.00***
Professional	-	-	-	-	-
Proprietor	-	-	-	-	-
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	0.89***	-	-
Age 16 to 25	-	-	-8.70***	-	-
Age 26 to 40	46.12*	-	-	-	-
	QR				
	.5	**.6**	**.7**	**.8**	**.9**
Intercept	-	-	-	-	-
Length of Stay	25.00*	43.20***	60.00***	84.07***	100.00***
Number of Visits	32.50***	39.20***	41.82***	48.98**	100.00***
Male	-	-	-23.64*	-	-
Married	-45.00*	-	-58.18*	-43.56*	-
Professional	-	-	-	-	-
Proprietor	-	-	68.18***	-	-
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	-	-	-
Age 26 to 40	-	23.20**	29.09***	50.00**	-

Figure 3: Least-squares and quantile regressions estimates for “Spending on Local Transportation.”



Note: Each dot represents a 5% quantile of the distribution. The solid curves that envelope the dots show the 95% confidence band for each 5% quantile dot and is significant when the "0" axis is outside of the envelope. The single thin dashed horizontal line is the least-squares mean of the entire distribution. The two dotted lines are the 95% confidence band for the entire distribution and is significant when the "0" axis is outside the two dotted lines. The 5% quantiles correspond directly to the 25% quartiles in Table 3.

Spending on Hotels

Wang (2004) found that the "length of stay" was not statistically significant in affecting spending on hotels for the *average* Mainland Chinese visitor to Hong Kong. Our least-squares result in Table 4 leads to the same conclusion. However, the quantile regression estimates in both Figure 4 (first row, first panel) and Table 4 show that the "length of stay" has a significant positive impact on hotel expenditure among the top 30% of the spenders in this category (the envelope lies above the horizontal axis for τ greater than 0.7). In fact, the top 10% are shown to spend as much as US$400 more for each additional day of stay compared to the roughly US$200 for the 0.7 quantile. The "number of visits" (first row, second panel), however, does not have any statistically significant effect on hotel spending for an *average* visitor (the two horizontal dotted lines enclose the horizontal axis). The quantile regression estimates also show no significant effect across the whole spectrum of the distribution (the entire envelope includes the horizontal axis).

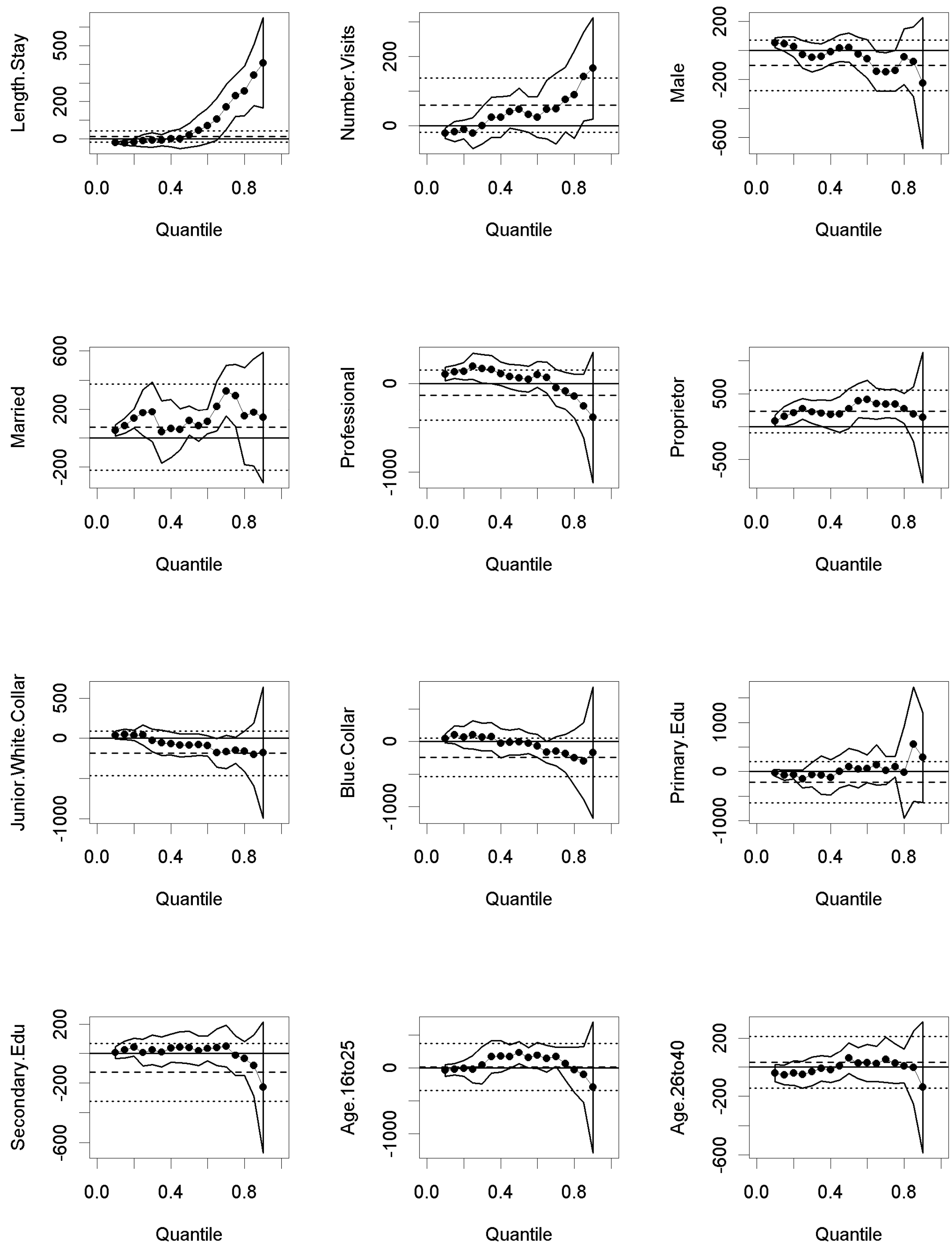
For the 0.1 to 0.2 and the 0.5 to 0.75 quantile cohort, being married (first row, last panel) leads to higher spending on hotels while there is no such effect for an *average* visitor. For example, the 10th percentile among the married tourists spend about US$50 more while the 70th percentile spend about US$325 more than their unmarried counterparts. Wang found that average "Proprietors/owners" spend about US$380 more than other job types while our least-squares estimate has found no such effect. This particular job type also spends more than the other job types on hotel for the 0.1 to 0.8 quantile with the exception of the 0.4 to 0.5 quantile.

The "age 16-25", "age 26-40" and "age 41-60" groups were found to spend significantly less than the base group in Wang (2004). However, we have found no such effect in our study (fourth row, second and third panels; "age 41-60" is not shown in Figure 4) for the confidence bands enclose the horizontal axis for all τ in both panel. We again attribute this to the different standard errors used in both studies.

Table 4. Estimated coefficients for the least-squares regression (LSR) and quantile regressions (QR) for "Spending on Hotel". The number of asterisks signifies the level at which the coefficients are significant with 0.05 (*), 0.01 (**) and 0.001 (***).

Variables	**LSR**	**QR**			
		.1	**.2**	**.3**	**.4**
Intercept	908.54***	102.17*	-	-	367.02**
Length of Stay	-	-19.57***	-	-	-
Number of Visits	-	-22.61**	-	-	-
Male	-	51.30**	-	-	-
Married	-	51.30**	135.22***	-	-
Professional	-	110.00**	141.30**	169.89*	-
Proprietor	-	85.22*	206.96*	224.71*	-
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	-	-	-
Age 26 to 40	-	-	-	-	-
	QR				
	.5	**.6**	**.7**	**.8**	**.9**
Intercept	287.50*	383.43*	-	-	-
Length of Stay	-	-	171.18**	255.77***	407.14***
Number of Visits	-	-	-	-	165.71*
Male	-	-	-150.00*	-	-
Married	120.00*	113.43*	325.88***	-	-
Professional	-	-	-	-	-
Proprietor	-	414.29**	347.06**	275.77*	-
Junior White Collar	-	-	-	-	-
Blue Collar	-	-	-	-	-
Primary /No Education	-	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	230.00**	-	167.06*	-	-
Age 26 to 40	-	-	-	-	-

Figure 4: Least-squares and quantile regressions estimates for “Spending on Hotels”.



Note: Each dot represents a 5% quantile of the distribution. The solid curves that envelope the dots show the 95% confidence band for each 5% quantile dot and is significant when the "0" axis is outside of the envelope. The single thin dashed horizontal line is the least-squares mean of the entire distribution. The two dotted lines are the 95% confidence band for the entire distribution and is significant when the "0" axis is outside the two dotted lines. The 5% quantiles correspond directly to the 25% quartiles in Table 4.

Total Spending

Although "length of stay" (Table 5 and Figure 5, first row, first panel) was found to positively impact "spending on meals outside hotel", "spending on local transportation" and "spending on hotels", in aggregate, it has no impact on total spending for an *average* Mainland Chinese visitor to Hong Kong, which is similar to Wang’s finding. It does, however, have a significant positive impact (about US$600) for the 0.7 to 0.8 quantile of these visitors.

Similar to Wang, the “number of visits” (Table 5 and Figure 5, first row, second panel) has a statistically significant positive impact on total spending for an *average* visitor. The quantile regression estimates, however, show that this positive effect is present only for the 0.3 quantile visitors. Again, when investigating the data more carefully, we discovered that this is caused by the extremely large outliers in total spending among those who have visited Hong Kong more than 3 times, which biases the least-squares estimate upward.

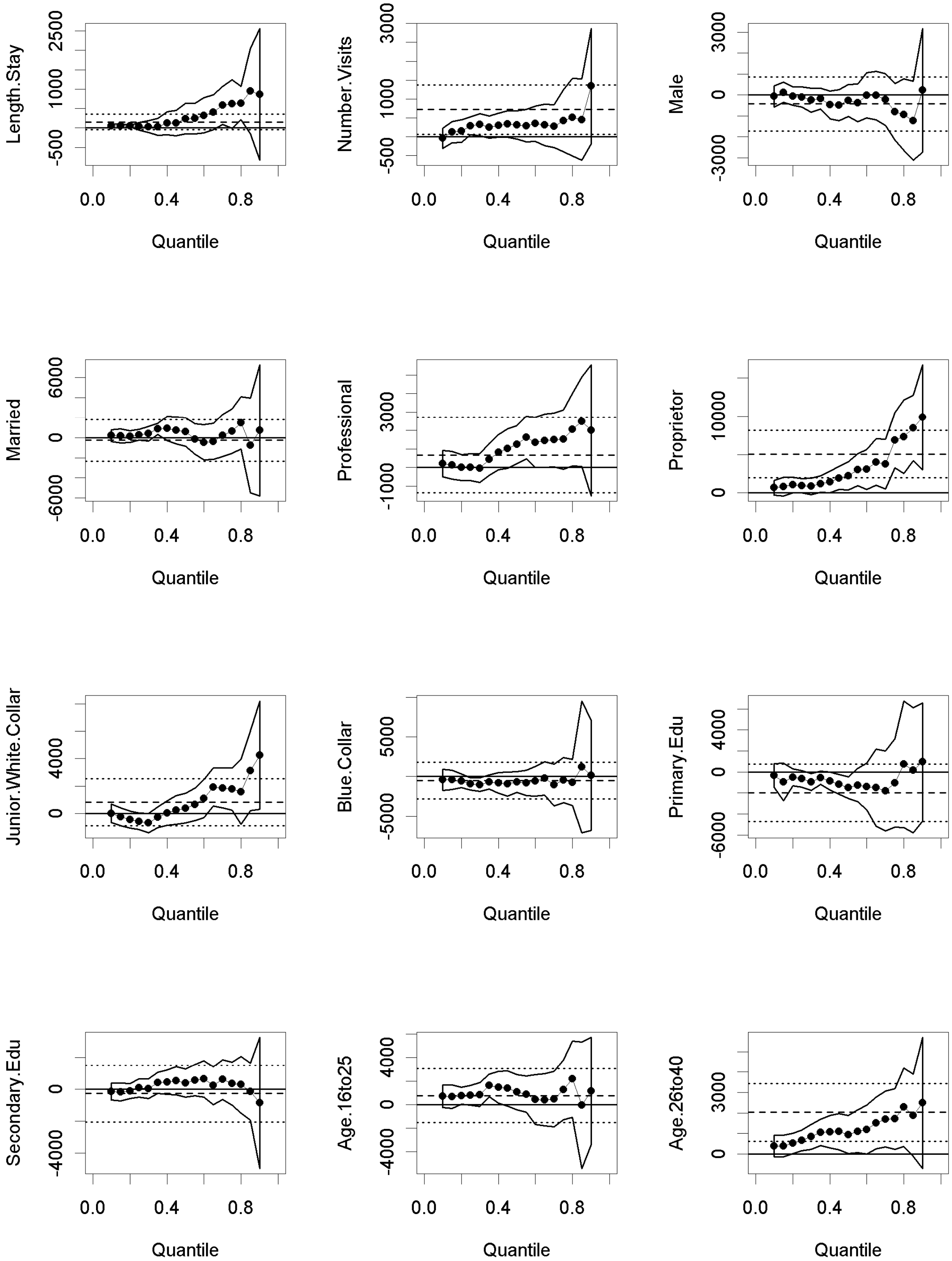
The *average* “professionals” do not have higher total spending according to the least-squares estimate in Table 5 and Figure 5 (second row, second panel), a result similar to Wang. But they do have higher total spending (around US$1,500) among the 0.5 to 0.8 quantile cohort when compared to the base level (“other job types”). Wang found that the "proprietors/owners" (second row, third panel) have a statistically significant higher total spending than the base level on *average* spender and so did we. This class of tourists also spend more for almost the whole spectrum of visitors except the 0.1, 0.15 and 0.3 quantile.

And finally, the “Age 26-40” group (fourth row, third panel) consistently spends more in total than the other age groups on *average*, which is contrary to Wang’s finding, and also across the whole spectrum of the distribution. The additional spending also increases as we move towards the upper tail of the distribution.

Table 5. Estimated coefficients for the least-squares regression (LSR) and quantile regressions (QR) for "Total Spending". The number of asterisks signifies the level at which the coefficients are significant with 0.05 (*), 0.01 (**) and 0.001 (***).

Variables	**LSR**	**QR**			
		.1	**.2**	**.3**	**.4**
Intercept	4037.30*	1087.04*	1837.90***	1997.23***	-
Length of Stay	-	56.52**	46.48***	-	-
Number of Visits	717.70*	-	-	322.00*	-
Male	-	-	-	-	-
Married	-	-	-	-	-
Professional	-	-	-	-	-
Proprietor	5054.00***	-	1065.41*	-	1431.17*
Junior White Collar	-	-	-	-705.62*	-
Blue Collar	-	-	-	-1020.81*	-
Primary /No Education	-	-	-	-939.19*	-869.92*
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	748.89*	-	1455.21*
Age 26 to 40	2025.90*	-	516.00*	834.81**	1083.25**
	QR				
	.5	**.6**	**.7**	**.8**	**.9**
Intercept	2080.67*	3344.80*	3110.11*	-	-
Length of Stay	-	-	581.14*	640.00**	-
Number of Visits	-	-	-	-	-
Male	-	-	-	-	-
Married	-	-	-	-	-
Professional	1237.22*	1360.80*	1494.97*	2047.82*	-
Proprietor	2232.17*	3060.40*	3748.86*	7320.73**	9875.18**
Junior White Collar	-	-	1864.57*	-	4260.00*
Blue Collar	-	-	-	-	-
Primary /No Education	-1496.83**	-	-	-	-
Secondary/High School	-	-	-	-	-
Age 16 to 25	-	-	-	-	-
Age 26 to 40	940.00*	-	1701.71*	2278.55*	-

Figure 5: Least-squares and quantile regressions estimates for “Total Spending”.



Note: Each dot represents a 5% quantile of the distribution. The solid curves that envelope the dots show the 95% confidence band for each 5% quantile dot and is significant when the "0" axis is outside of the envelope. The single thin dashed horizontal line is the least-squares mean of the entire distribution. The two dotted lines are the 95% confidence band for the entire distribution and is significant when the "0" axis is outside the two dotted lines. The 5% quantiles correspond directly to the 25% quartiles in Table 5.

Discussions and Conclusions

Some of the least-squares regression results in this study are the same as those found in Wang (2004). For the *average* Mainland Chinese visitor to Hong Kong, we also found that "length of stay" is statistically significant at the 5% level in affecting "spending on local transportation" for an *average* visitor. The "number of visits" has a significant impact on "spending on meals" and "spending on local transportation", and "total spending", as Wang found. And "proprietors/owners" spend on *average* more than the other job types on shopping and meals outside hotel.

There are also significant differences between the two studies. Contrary to Wang (2004), we cannot conclude that the "number of visits" has a statistically significant impact on "spending for shopping", nor can we conclude that the "length of stay" is statistically significant for "spending for meals outside hotel" for the *average* visitor using the heteroskedasticity-robust standard errors for the least-squares estimates in light of the presence of heteroskedasticity in the models. Neither can we conclude that "proprietors/owners" spend on *average* more than the other job types on "hotels", or that any of the age groups spend more on hotels. On the contrary, however, we have found that the *average* visitor among the "age 26 to 40" group has higher spending on "shopping", "local transportation", and "total spending" when compared to the other age groups.

More importantly, we have unveiled, through the use of quantile regressions, additional interesting effects of the covariates on the various categories of consumptions that were not revealed in Wang (2004). Even though the "length of stay" is not significant in affecting "spending on meals outside hotel" for an *average* visitor, it has a positive impact for the upper 50% of visitor spending in this category. The longer the upper half of the visitors stay, they will spend increasingly more on meals outside hotel. This "length of stay" effect is five times larger for the top 10th percentile than the median. This is good news for higher-end restaurants. They may be able to increase their revenue strategically by participating in programs that encourage longer stays among this half of "heavier" spenders on meals. Besides having a positive effect for the *average* visitor, the "length of stay" also has significant positive influences on "local transportation" spending throughout the whole spectrum of the distribution, as well as a positive impact on "spending on hotels" for the top 30th percentiles. High-end hotel chains that want to increase their revenue would see their biggest impact from a focus on the "heavier spenders" who comprise the top 30% of current spenders on hotels.

Though the "number of visits" is not significant in influencing "spending for shopping" for an *average* visitor, it does have a significant impact for the middle 50% of the spenders on shopping. So tourism boards, cities or any municipal entities might want to target the middle 50% of repeat visitors for shopping purposes. Higher number of visits also results in higher spending on local transportation and this higher amount increases even more as we move from the bottom to the top of the distribution.

The top 40% of the Mainland Chinese shoppers to Hong Kong who are "proprietors/owners" spend more than the other job types and the top 10% of them spend over four times more than the 60th percentile shoppers. Again organizations that promote shopping might want to target the higher spenders among the "proprietors/owners". The middle 30% of the "proprietors/owners" category spend more on meals outside hotel. "Proprietors/owners" also spend more than other job types on "hotels" throughout almost the whole distribution. In aggregate, this job type has higher "total spending" throughout almost the whole distribution spectrum and the amount of higher spending increases as we move from the bottom towards the top of the distribution. The tourism board in Hong Kong might want to especially focus on this particular occupational group.

The "age 26 to 40" group spends significantly more than the other age groups on shopping over the whole distribution, and the incremental spending is higher for the higher quantiles. This suggests that shops might want to target this age group of shoppers with their advertising campaign. This group also has higher spending on "local transportation" for the 0.6 to 0.8 quantiles. In aggregate, they also have higher "total spending "than the other age groups.

We have seen that different spending cohorts (defined by the conditional quantiles of the various spending categories) of visitors, with the same socio-economical characteristics, length of stay and number of visits, can have very different consumption behaviors. Understanding these differences in spending behaviors can be very useful for restaurant owners, hotel chains, transportation providers and shopping centers when allocating their limited resources.

We should emphasize that the findings in this study are drawn from data collected from China's mainland visitors to Hong Kong. Caution should be exercised when drawing implications for visitors from other origins and to other tourist destinations. For example, the higher spending on shopping among the "proprietors/owners" may be due to the higher purchasing power of the newly emerging class of self-proprietors/owners encouraged by the economic reform policy over the past 30 years in China. This high purchasing power might not exist among visitors to other tourist destinations who are "proprietors/owners". Similarly, the higher shopping expenditure among the "26 to 40 age group" might have reflected the exceptionally high demand for luxurious goods available only in Hong Kong fueled by high purchasing power among this group of visitors. In other tourist destinations that attract visitors from countries where luxurious goods are more readily available, we may not see the same higher shopping appetite among the same age group. Furthermore, as with all survey data, these results represent a snapshot of tourist behavior during a specific time period and under macroeconomic conditions that applied at that time.

A significant consideration in this and any study of tourist expenditures is the accuracy of respondent recall. Although Frechtling (1994b: 368) reviewed eight approaches to assessing tourist expenditures, including direct observation, residual receipts, and cost factors, by far the dominant approach is that of sample surveys. Tourism expenditure surveys can occur before (projected estimates), during or after the trip, and can include considerable detail on trip characteristics, respondent demographics and expenditure types. Daily diaries are the most accurate approach, though response rates are the lowest due to the time requirements (Burke and Gitelson 1990). All surveys, however, suffer from tourist recall bias and a generally under-reporting expenditures in the short term, moving towards over-reporting them in the longer term (Mak et al 1977). Tourist surveys also suffer from a lack among most respondents as to the proportion of the full or partial travel package they purchased that should be allocated to accommodations, transportation, meals, tours and travel agents. Response rates and the validity of the sample set are further considerations affecting the accuracy of the results.

One other method of analysis has been proposed in recent years to help destinations better distinguish "big spenders" from other visitors. This is the CHAID methodology, used by Legohérel and Wong (2006) and Díaz-Pérez, et al. (2005). The CHAID methodology uses a decision tree to identify the variables that most differences in tourist segments, with each level of the tree adding a new significance factor. For example, Díaz-Pérez, et al. (2005) found nationality to be the first distinguishing factor, with the star level of the accommodation used was the second level. The applicable varible third level varied among each the second level groups, though occupation was most significant overall. Thus, the combination of these three variable was able to explain most of the difference in tourist expenditures, and could be used to develop marketing plans for different segments. The CHAID approach is effective in identifying market categories that may not be obvious in traditional market analysis and anecdotal assumptions. This complements nicely with the quantile regression technique that we have introduced in this study. The quantile regression analysis is able to discover different marginal effects that are not obvious. In particular, the quantile regression shows the significance of each independent variable on the dependent variable (expenditure) across the full spectrum of the population distribution. For example, we have shown that an additional day of stay has a higher impact on spending on meals for the upper 50% heavier spenders than it does for the lower half. The CHAID approach can then be used to help identify the heavy spenders based on their socio-economic characteristics.

Acknowledgement

The authors wish to thank Dr. Donggen Wang , Department of Geography, Hong Kong Baptist University, for the use of his original data set for this study. This study was partially funded by the U.S. Fulbright Scholarship program.

References

Agarwal, V.B. and Yochum, G.R. 1999. Tourist Spending and Race of Visitors. *Journal of Travel Research 38(2): 173-176.*

Barrodale, I. and F. D. K. Roberts (1974), "Solution of an overdetermined system of equation in the l_1 norm", Communications of the ACM, 17(6), 319-320.

Burke, J.F. and Gitelson, R. 1990. Conversion Studies: Assumpptions, applications, accuracy and abuse. *Journal of Travel Research* 27 (Winter): 46-51.

Charles, K. K., E. Hurst and N. Roussanov (2007), "Conspicuous Consumption and Race", URL http://knowledge.wharton.upenn.edu/papers/1353.pdf.

Díaz-Pérez, F.M.; Bethencourt-Cejas, M; and Álvarez-González, J.A.A. (2005) The segmentation of canary island tourism markets by expenditure: implications for tourism policy. Tourism Management 26:961-964.

English, D.B.K. 2000. Calculating Confidence Intervals for Regional Economic Impacts of Recreation by Bootstrapping Visitor Expenditures. Jounral of Regional Science 40(3): 523-539.

Frechtling, D.C. 1994a. Assessing the Economic Impacts of Travel and Tourism -- Introduction to Travel Economic Impact Estimation. In J.R.B. Ritchie and C.R. Goeldner, eds., *Travel, Tourism, and Hospitality Research: A Handbook for Managers and Researchers*, pp. 359-365. New York: Wiley.

Frechtling, D.C. 1994b. Assessing the Impacts of Travel and Tourism -- Measuring Economic Benefits. In J.R.B. Ritchie and C.R. Goeldner, eds., *Travel, Tourism, and Hospitality Research: A Handbook for Managers and Researchers*, pp. 367-391. New York: Wiley.

Hall, C.M. and A.A. Lew (2009) *Understanding and Managing Tourism Impacts: An Integrated Approach*. Oxford, UK: Routledge. (forthcoming)

Heung, V. C. S. and Cheng, E. (2000) 'Assessing Tourists' Satisfaction with Shopping in the Hong Kong Special Administrative Administrative Region of China', *Journal of Travel Research*, 38(4): 396–404.

Joppe, M.; Martin, D.W. and Waalen, J. (2001) Toronto's Image As a Destination: A Comparative Importance-Satisfaction Analysis by Origin of Visitor. *Journal of Travel Research* 39 (3):252-260.

Koenker, R. (2009), *quantreg: Quantile Regression*, R package version 4.26, http://www.r-project.org.

Koenker, R. and V. d'Orey (1987), "Algorithm AS 229: Computing Regression Quantiles", *Applied Statistics*, 36(3), 383-393.

Koenker, R. and V. d'Orey (1994), "Remark AS R92: A Remark on Algorithm AS 229: Computing Dual Regression Quantiles and Regression Rank Scores", *Applied Statistics*, 43(2), 410-414.

Legohérel, P. and Wong, K.K.F. 2006. Market Segmentation in the Tourism Industry and Consumers' Spending: What About Direct Expenditures? *Journal of Travel & Tourism Marketing*, Vol. 20(2): 15-30.

Lehto, X.Y.; Cai, L.A.; O'Leary, J.T. and Huan, T-C. (2004) Tourist shopping preferences and expenditure behaviours: The case of the Taiwanese outbound market. *Journal of Vacation Marketing* 10(4):320-332.

Mak, J., Moncur, J. and Yonamine, D. 1977. How or How Not to Measure Visitor Expenditure. *Journal of Travel Research* 16 (Summer): 1-4.

Mehmetoglu, M. (2007) Nature-based Tourists: The Relationship Between their Trip Expenditures and Activities. *Journal of Sustainable Tourism* 15(2): 200-215.

Mok, C. and Iverson, T.J. (2000) Expenditure-based segmentation: Taiwanese tourists to Guam. *Tourism Management* 21: 299-305.

Narayan, P.K. (2005) The structure of tourist expenditure in Fiji: evidence from unit root structural break tests. *Applied Economics* 37: 1157–1161

Oh, J.Y-J.; Cheng, C.K.; Lehto, X.Y. and O'Leary, J.T. (2004) Predictors of tourists' shopping behaviour: Examination of socio-demographic characteristics and trip typologies. *Journal of Vacation Marketing* 10(4): 308-319.

R Development Core Team (2008), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Suh, K.S. and Gartner, W.C. (2004) Preferences and trip expenditures—a conjoint analysis of visitors to Seoul, Korea. *Tourism Management* 25: 127-137.

Wang, D. (2004), "Tourist Behaviour and Repeat Visitation to Hong Kong", *Tourism Geographie*, 6 (1), 99–118.

Weber, S. (1995). Psychographic segmentation. In S.F. Witt & L. Moutinho (Eds.), *Tourism marketing and management handbook* (pp. 316-324). Hemel Hempstead: Prentice Hall.

Wicks, B. E., & Schuett, M. A. (1993). Using travel brochures to target frequent travellers and "big-spenders." *Journal of Travel and Tourism Marketing*, 2(2/3), 77-90.

Wong, J. and Law, R. (2003) 'Differences in Shopping Satisfaction Levels: A Study of Tourists in Hong Kong', *Tourism Management*, 32(4): 401–10