Analysis of Domestic Tourist Expenditure in Colombia Using a Quantile Selection Approach

Oscar Jaulín-Méndez*, Andrey Ramos†

March, 2023

Abstract

This paper investigates the determinants of domestic tourist expenditure in Colombia through the application of both linear and quantile regression models correcting for sample selection. The methodological approach enables us to examine the potential heterogeneity in the effects of a set of explanatory variables across the conditional distribution of tourist expenditure. Income, individual characteristics, and trip-related factors such as motivation and destination are found to be relevant in driving domestic tourists' decisions. Furthermore, these variables exhibit heterogeneous effects depending on the level of tourist expenditure. For those tourist with high level of expenditure, income and education emerge are relevant determinants; in contrast, lower quantiles of the distribution are more affected by age or the presence of children, variables linked to time and preferences restrictions.

Keywords: Tourist expenditure; Domestic tourism; Sample selection; Quantile regression; Colombia.

JEL Classification: C31, L83, Z31.

1 Introduction

The tourism sector is recognized as a source of economic growth and employment in modern economies (Brida et al., 2016; Song et al., 2012). One of the primary mechanisms through which destination communities benefit from tourism flows is

*Department of Economics, Universidad Carlos III de Madrid. osjaulin@eco.uc3m.es
†Department of Economics, Universidad Carlos III de Madrid. anramosr@eco.uc3m.es

through tourist expenditure in various categories, including transportation, lodging, food, attraction entrances, and souvenirs (Brida and Scuderi, 2013; Figini and Vici, 2010). For both destination managers and policymakers, having access to precise and reliable information about the factors that influence tourist expenditure is crucial for designing policies aimed at maximizing the economic revenue of tourism. To serve this purpose, a substantial body of literature has explored the determinants of tourist expenditure in various contexts (Brida et al., 2018; Park et al., 2020; Thrane, 2016).

In this paper, we focus on analyzing the determinants of domestic tourist expenditures in Colombia using data from the *Encuesta de Gasto Interno en Turismo* (EGIT) conducted by the *Departamento Administrativo Nacional de Estadística* (DANE) in 2019. These determinants includes sociodemographic variables such as income, gender, age, educational level, and occupation which account for economic, time-related, preference-related, and cultural factors influencing both travel decisions and spending levels. Additionally we include trip-related factors as motivation or destination.

Most of the existing empirical literature employs linear regression methods to focus on the average response of tourism expenditure to various potential determinants. However, in this study, we expand the scope of our analysis by estimating both conditional mean and quantile regression models. Unlike conditional mean models, the conditional quantile regression model allows us to investigate the heterogeneous effects of the explanatory variables across the entire distribution of tourist expenditure, as demonstrated in previous works such as Lew and Ng (2012), Marrocu et al. (2015) and Mitra et al. (2019). Notably, these prior studies did not address the potential presence of sample selection bias in their estimations. To address this issue, we employ the correction method proposed by Arellano and Bonhomme (2017) to obtain consistent estimators.

The rest of the document is organized as follows. Related literature review is presented in Section 2. Section 3 outlines the econometric methodology. Section 4 offers an explanation of the data and variables used in the empirical analysis. The results are described in Section 5. Section 6 concludes and briefly discuss some potential paths for future research.

2 Related literature

The determinants of tourist expenditure have been extensively studied from both aggregate and individual perspectives. For a comprehensive review of econometric approaches for analyzing tourism expenditure at the individual level, please refer to Brida and Scuderi (2013) and the references therein. The literature primarily employs two main econometric approaches to identify the individual determinants

of tourist spending. On one hand, some studies use linear regressions, both with and without corrections for sample selection bias, to investigate the impact of a set of explanatory variables on the mean of the tourist spending. On the other hand, other studies employ conditional and unconditional quantile regression approaches to examine the determinants across different segments of the expenditure distribution. Nonetheless, they do not incorporate methodologies to correct for sample selection bias.

According to Wang and Davidson (2010), the examination of the determinants of the mean tourist expenditure began with the early work of Mak et al. (1977). However, this topic gained more attention in the literature towards the end of the 1990s decade and the beginning of the 21st century. In this initial phase of analysis, most studies relied on Ordinary Least Squares (OLS) regression techniques. They found that economic variables, such as income and price, are important drivers of tourist expenditure behavior. In contrast, the conclusions regarding socio-demographic variables as age, gender, or education, remained mixed (Wang and Davidson, 2010). Some notable examples of this type of literature include Asgary et al. (1997), Wang et al. (2006), Cannon and Ford (2002), Seiler et al. (2003), Wang and Davidson (2010), and Thrane and Farstad (2011), among others. Additionally, some documents adopted the Tobit model as an estimation strategy, particularly when the independent variable was derived from survey data with censored observations (Barquet et al., 2011; Kim et al., 2011; Zheng and Zhang, 2013).

None of the aforementioned studies faced the potential sample selection bias that can arise when estimating the impact of explanatory variables on expenditure. This bias emerges because the sampling procedure is not random and is driven by observed and/or unobserved characteristics. For instance, the decision to travel may be influenced by the values of certain characteristics, rendering the sample of travelers non-random. Since expenditure is incurred only by tourists, achieving consistent parameter estimators necessitates the correction for selection bias in the travelling decision. The most widely employed approach to address this bias is the one proposed by Heckman (1979). In the literature on tourist spending, this approach is applied and discussed in studies such as Jang and Ham (2009), Brida et al. (2014), and Thrane (2016), among others.

Another branch of research aims to identify the determinants of tourist spending across its distribution rather than solely focusing on its average. These studies allow for the exploration of the heterogeneity in responses at different spending levels. Within this type of literature, we find two alternative approaches: one based on the conditional quantile regression method proposed by Koenker and Bassett (1978) and the other based on the unconditional quantile regression tech-

nique introduced by Firpo et al. (2009).[1] Both approaches provide evidence of heterogeneous responses. Applications of the conditional quantile regression include Lew and Ng (2012), Marrocu et al. (2015), and Mitra et al. (2019), while studies employing the alternative unconditional are Sharma et al. (2020), Pérez-Rodríguez and Ledesma-Rodriguez (2021), Azam (2022), and Sahoo et al. (2022). Notably, these studies do not explicitly address the issue of sample selection bias in empirical estimators. Arellano and Bonhomme (2017) proposed a methodology to deal with sample selection in conditional quantile regression that has not been used yet in this literature.

The contribution of this paper to the existing literature is threefold. First, it investigates the determinants of tourist expenditure in Colombia, an emerging economy with a growing tourism sector,[2] but with surprisingly few applications in this field.We highligh the work by Alzate and Espinal (2018) and Góngora and Osorio (2020) who analyze tourism decisions in Colombia using data from the same source. However, Alzate and Espinal (2018) focuses on characterizing the origin-destination matrix of tourism, while Góngora and Osorio (2020) examines the determinants of the decision to travel rather than the level of expenditure. Second, our study models domestic tourist expenditure, which refers to the spending made by Colombian citizens when traveling within the country. With a few exceptions, such as Thrane and Farstad (2011), Thrane (2016), or Bel et al. (2015), most applied research focuses on expenditure by foreign tourists. Third, to the best of our knowledge, this paper represents the first attempt to analyze tourist expenditure while correcting for sample selection within a quantile regression framework, employing the approach proposed by Arellano and Bonhomme (2017).

3 Econometric Methodology

The determinants of the domestic tourist expenditures in Colombia are analyzed using conditional mean and conditional quantile regression models. The former approach focuses on estimating the average response of domestic tourist expenditure to changes in a set of explanatory variables. This analysis is enhanced by exploring specific segments of the expenditure's conditional distribution through quantile regressions. This approach provides a more comprehensive understanding of spending behavior, as it allows us to examine how the effects of each determinant vary across different spending levels.

The estimates obtained from the direct application of Ordinary Least Squares

[1] See Borah and Basu (2013) for a complete discussion on the differences between conditional and unconditional quantile regression approaches.

[2] According to DANE, participation of lodging and food services on real GDP increased 26.0 basic points (bp) from 2005 to 2019.

(OLS) and conditional quantile regressions (see Koenker and Bassett (1978) for more details) may suffer from sample selection bias. Specifically, respondents in the EGIT only answer questions related to expenditure decisions if they choose to travel within the country for tourism purposes. Given that the decision to travel is not random and may be correlated with spending behavior, the group of domestic tourists may differ systematically from non-tourist citizens. The econometrics literature documents the empirical implications of sample selection bias (see Heckman (2010) and references therein). Consequently, it becomes necessary to apply correction techniques to obtain consistent parameter estimates.

In the case of conditional mean regression, we employ the well-known approach proposed by Heckman (1979) to correct for sample selection bias. In the context of conditional quantile regressions, we apply the methodology introduced by Arellano and Bonhomme (2017). This correction process involves a three-step procedure as follows. First, a consistent estimator of the propensity score parameter ($\hat{\theta}$) is obtained by utilizing a probit model for the selection equation. In this step, the dependent variable is an indicator variable, taking a value of 1 if the individual is a domestic tourist and 0 otherwise. In the second step, we estimate a copula parameter ($\hat{\rho}$) that quantifies the dependence between the errors in the equation for the tourist expenditure (or outcome equation) and the selection equation. This estimation minimizes an objective function based on the method of moments, which considers the Joint CDF of the percentile error in the outcome equation and the error in the selection equation. Finally, a consistent estimator of the τ-th quantile regression coefficient is obtained as:

$$\hat{\beta}_\tau = \underset{b \in B}{\operatorname{argmin}} \sum_{i=1}^{N} D_i[\hat{G}_{\tau i}(Y_i - X_i' b^+) + (1 - \hat{G}_{\tau i})(Y_i - X_i' b^-)] \tag{1}$$

where D_i is the selection indicator, B is the parameter space for $\hat{\beta}_\tau$, $a^+ = max(0, a)$, $a^- = max(-a, 0)$, and

$$\hat{G}_{\tau i} = G(\tau, \rho(Z_i; \hat{\theta}); \hat{\rho}) \tag{2}$$

where $\hat{\theta}$ is a consistent estimate of the propensity score parameter obtained in the first step, Z_i is the vector of explanatory variables included in the selection equation, $X_i \subset Z_i$ is the vector of explanatory variables included in the dependent variable equation and $G(.)$ is the copula function.

The copula of a normal bivariate distribution is considered (which depends on the value of correlation between variables, ρ) and the instrument function is the propensity score. The selection and outcome equations include explanatory variables, which will be detailed in the following section. These variables were selected based on the existing literature in the topic and data availability. The parameters of interest are estimated following Biewen and Erhardt (2021). Standard errors are computed using Bootstrap with 1000 draws.

3.1 Exclusion restriction

Sample selection correction requires imposing an exclusion restriction to identify the model parameters. In our context, we consider as exclusion restriction the quarter in which the individual participated in the EGIT. We argue that tourism is a seasonal activity that in Colombia increases in December and July. Since EGIT classifies an individual as a tourist if she traveled in the previous month to the survey, we claim that individuals who answered the survey in the first (January-March) or third (July-September) quarter are more likely to travel than individuals interviewed in other quarters. Because individuals cannot select the quarter in which to respond to the EGIT, the variable is considered exogenous and does not effect on spending behavior.

4 Data and Variables

The data for this study are sourced from the EGIT conducted by the DANE in 2019. This survey captures information regarding travel decisions made by individuals residing in Colombia's 24 major cities. It collects data on various aspects, including travel decisions, tourist expenditures, travel motivations, and destinations. Additionally, it includes sociodemographic variables such as income, gender, age, educational level, and occupation, that are necessary to account for economic, time, preference, and cultural-related factors influencing both travel decisions and spending levels (Brida and Scuderi, 2013; Marrocu et al., 2015).

In our empirical analysis, we utilize data provided by the head of the household to avoid potential overestimation of spending at the household level. Our variable of interest is defined as the natural logarithm of daily per-person tourist expenditure. While the empirical literature often considers variables like length of stay and travel group size as independent variables, we normalize expenditure by these variables to mitigate division bias, as recommended by Borjas (1980).

According to the EGIT data, from a total sample of 56,381 heads of household, less than 10% (5,313) are classified as domestic tourists. Samples of tourists and non-tourists are heterogeneous in terms of observable characteristics. Table 1 shows that individuals in the sample of tourists have in average a higher income, whereas the non-tourist sample is composed by a higher proportion of individuals living in low socioeconomic strata and with low levels of education. The imbalance in the referred characteristics is a first indicator of sample selection.

We compute the histogram and box plot for the expenditure variable using the sample of domestic tourists, as shown in Figure 1. It is evident that the distribution of tourist spending is asymmetric and concentrated in lower values. The average daily spending per person is estimated at $74,725.34$ Colombian pesos

(COP), which is approximately 23 US Dollars (USD), while the median is smaller at around 44,444.45 COP, roughly 14 USD.[3] Notably, the box plot reveals the presence of outliers in the right tail of the distribution. The primary motives for travel are visiting relatives (49%) or vacationing (29%), while the Caribbean region (25%) and the Pacific region (16%) are among the most preferred destinations. This suggests a predominant trend of sun and beach domestic tourism, see Figure 2.

Table 1: Descriptive statistics by sample

Variable	Full	Non-Tourists	Tourists	Diff
Sociodemographic				
Log-income	13.706	13.665	14.076	-0.411***
Age	50.286	50.641	47.082	3.559***
Male	0.562	0.561	0.572	0.1094
Strata 0	0.003	0.003	0.001	0.002***
Strata 1	0.272	0.284	0.161	0.123***
Strata 2	0.359	0.367	0.295	0.072***
Strata 3	0.259	0.251	0.334	-0.083***
Strata 4	0.071	0.064	0.126	-0.062***
Strata 5	0.022	0.019	0.048	-0.029***
Strata 6	0.013	0.011	0.034	-0.023***
Children	0.402	0.411	0.316	0.095***
Recreational house	0.016	0.012	0.057	-0.045***
Education				
None	0.034	0.037	0.012	0.025***
Primary	0.233	0.244	0.128	0.116***
Secondary	0.128	0.134	0.0815	0.052***
Upper secondary	0.285	0.290	0.234	0.056***
Superior	0.319	0.295	0.544	-0.249***

Notes: The table contains the mean of the Sociodemographic and Educational variables for the Full sample (61,694 obs.), as well as for the non-tourist (56,381 obs.) and tourist (5,313 obs.) samples. Additionally the results of the test for the null of no mean differences between tourist and non-tourist are reported. In such a case, p-values are represented as *** p<0.01, ** p<0.05, * p<0.1.

[3] As of December 31, 2019, 1 USD was equivalent to 3,285.76 COP. The minimum wage in Colombia in 2019 was 828,116 COP, equivalent to 223 USD.

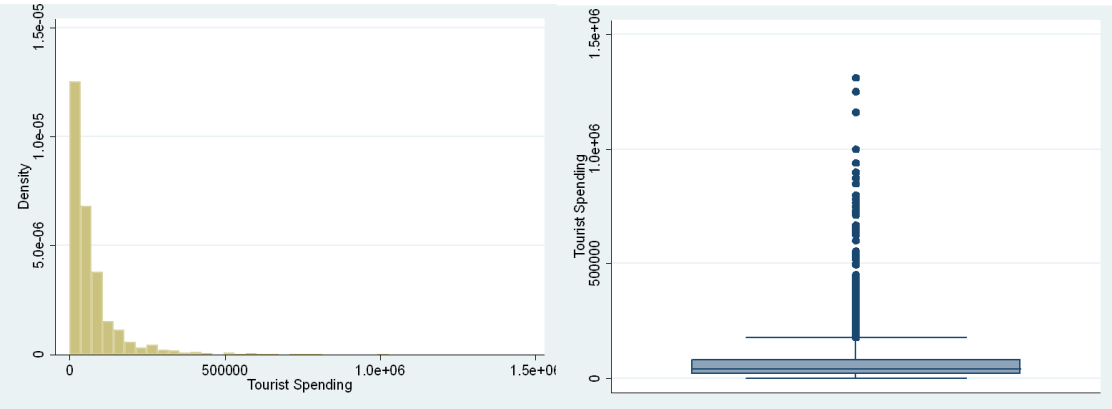
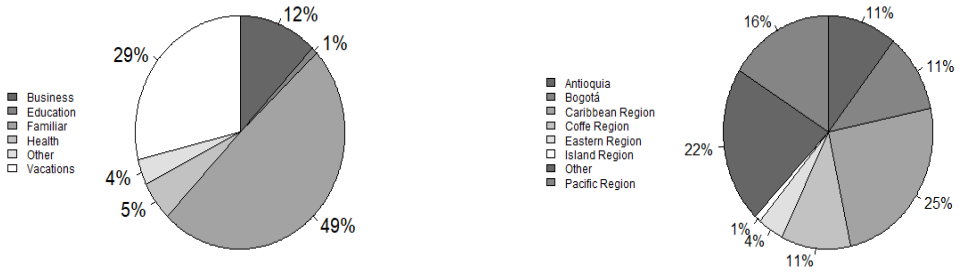


Figure 1: Histogram and box plot of tourist spending



(a) Motive

(b) Destination

Figure 2: Tourists by motive and destination.

5 Results

In this section, we present the results obtained from applying both linear and quantile regression models correcting for sample selection. The dependent variable in the selection equation corresponds to the travel decision, while the dependent variable in the outcome equation represents the logarithm of daily per-person tourist expenditure. Region of residence, marital status, and occupation are included as control variables in all regressions, although we do not report estimates for these variables.

5.1 Conditional-mean regression correcting for sample selection

In Table 2, we present the estimates obtained by fitting a linear regression model to the tourist expenditure variable, following the approach of Heckman (1979) to correct for sample selection bias. It is important to note that most regressors (except for income and age) are binary variables, and the coefficients should be interpreted as the differential effect relative to the reference category. The statistical significance of the coefficients is assessed based on heteroskedasticity-robust standard errors.

The first column in Table 2 reports the estimates of the selection equation, which helps to establish the variables explaining the decision to travel for tourism motives in Colombia. Notably, income, gender, and age emerge as significant factors in determining travel choices: older individuals, males, and those with higher income are more likely to travel. Variables such as economic strata and education also positively impact the probability of being a tourist, with this effect increasing monotonically as these variables' levels rise. Conversely, having children in the household reduces the likelihood of traveling, possibly due to time constraints. Similar results are obtained in Góngora and Osorio (2020) and point to the existence of income, time, preference, and cultural restrictions that determine travelling decisions in Colombia.

The statistical significance of the survey quarter indicates support for the exclusion restriction assumed in our identification strategy. The dependence parameter (ρ) is statistically significant, indicating the presence of sample selection bias. The corrected estimates in the second column of Table 2 are informative about the expected response of the conditional mean of tourist expenditure to changes in specific determinants, *ceteris paribus*. Our findings align with existing evidence regarding the significance of income as a driver of tourist expenditure. In particular, the income-elasticity coefficient is estimated at 0.03318, and it is statistically significant at the 1% level.

Other sociodemographic variables are also relevant to explain tourist expenditures. Age, gender, education, strata, and having children are statistically significant, with expected signs and magnitudes. Notably, the association between age and tourist expenditure exhibits a traditionally U-inverted shape, indicating that tourist expenditure increases with age but begins to decrease after reaching a certain threshold. Our findings regarding the non-linear relationship between age and tourist expenditure align with similar results reported by Thrane and Farstad (2011). The estimated coefficients for education and strata monotonically increase with the levels of these variables, mirroring observations in the selection equation. This finding may be associated with lower tourism information-related search costs for individuals with higher education levels, providing them with more

opportunities to travel. Additionally, their enhanced communicative competence, knowledge, and cultural capital accumulation may contribute to higher spending, as suggested by Hung et al. (2012).

Turning to trip-related variables, both motivation and destination significantly influence tourist expenditure. When compared to travelers motivated by business purposes, those traveling for vacation, family visits, health-related reasons, or religious purposes tend to spend less on average. In contrast, tourists motivated by shopping tend to spend more, as expected. Regarding destination, tourists in all regions except the Island region exhibit lower average expenditure compared to those travelling to Bogotá. This result is not surprising in the context of Colombia. The Island region, situated in the Caribbean Sea near the coast of Nicaragua, requires tourists to incur additional expenses such as plane tickets or tourist packages for hotels, which can lead to higher spending compared to other destinations.

Overall, our findings from the conditional mean regression align the main results of previous empirical literature on tourist expenditure (see Brida et al. (2016)). While we have accounted for both demand and supply-related heterogeneity through a comprehensive set of explanatory variables, it's possible that heterogeneity is also reflected in the coefficients, as the same determinant may yield different effects depending on the level of tourist expenditure. This is particularly important in this context because the distribution of tourist spending is not symmetric and outliers are present in the data, therefore, the average is not totally informative about the relationship between tourist spending and its determinants. We investigate this issue further in the next section using conditional quantile regressions.

5.2 Conditional-quantile regressions correcting for sample selection

The econometric estimates, obtained using the approach of Arellano and Bonhomme (2017) to correct for sample selection bias in a quantile-regression framework, are presented in Table 3. The estimates of the selection equation are similar to those presented in the previous section. Small changes arise because of differences in estimation methods, but the conclusions remain unaltered.[4] Conditional quantile regression analysis was performed for selected percentiles of the tourist expenditure distribution, namely $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$.

Turning to the estimates obtained for different quantile regressions, these coef-

[4]The main difference is the estimation of the dependence parameter ρ. While Heckman (1979) uses a full information Maximum Likelihood approach, Arellano and Bonhomme (2017) use a method of moments estimator explained below. Although there exist such difference, there is evidence of correlation among errors in the selection and observation equations in both strategies.

Table 2: Conditional mean regression estimates correcting for sample selection

Variables	Heckman estimates	
	Selection Equation	Outcome equation
A. Sociodemographic		
Log-income	0.0404***	0.0331***
Age	0.0072**	0.0218***
Age^2	-0.0001***	-0.00021***
Male	0.0390**	0.1886***
Strata 1	0.2239	0.3535
Strata 2	0.3042	0.5769*
Strata 3	0.4804**	0.7364**
Strata 4	0.6451***	.9174**
Strata 5	0.6726***	1.1448***
Strata 6	0.7505***	1.1995***
Children	-0.0799***	-0.1477***
Recreational house	0.6454***	-0.04489
B. Education (Ref: None)		
Primary	0.1050*	0.1432
Secondary	0.1252*	0.2335**
Upper secondary	0.2138***	0.3609***
Superior	0.4969***	0.6159***
C. Motive (Ref: Business)		
Vacation		-0.1911**
Familiar		-0.6500***
Education		0.1346
Health		-0.1870***
Religion		-0.3443***
Shopping		1.140 ***
Other		-0.11508
D. Destination (Ref: Bogotá)		
Coffee region		-0.6094***
Caribbean region		-0.190***
Island region		0.9572***
Pacific region		-0.5490***
Antioquia		-0.0741
Eastern region		-0.5447**
Other		-0.4888 ***
E. Exclusion restriction		
Quarter 2	-0.0428**	
Quarter 3	-0.1050***	
Quarter 4	-0.1508***	
F. Dependence parameter		
ρ	0.3944**	

Notes: The table contains the results of the regressions correcting for sample selection. Individuals whose answer to strata and education questions was "not knowing" were excluded from the sample. Individuals with a daily per-person expenditure smaller than 0,30 U.S dollars were treated as non-travelers those. Individuals with kindergarden education and no education were included in the same group. In all the regressions we control for Region of residence, marital status and occupation. p-values for significance tests are represented as *** p<0.01, ** p<0.05, * p<0.1.

ficients are informative about the effect of explanatory variables on the conditional distribution of tourist expenditure. Notably, they reveal significant implications both at the median ($\tau = 0.5$) and across the complete expenditure distribution. Focusing on the median regression, it's observed that tourist expenditure exhibits a significant income elasticity of approximately 0.0221, which is smaller compared to the estimates for the mean. The difference in the estimates are attributed to the presence of outliers and the asymmetric nature of the tourist expenditure distribution, as described in Figure 1. Age and strata do not exhibit a statistically significant effect, although their point estimates show the expected positive direction.

Among sociodemographic variables, only higher education levels seem to have a significant effect on this percentile. Trip-related variables, such as motivation and destination, also influence expenditure levels, consistent with prior literature and the results obtained for the mean regression.[5] Regarding motives, shopping tourists exhibit a positive and significant marginal effect compared to the reference category (Business), while other motivations have a negative effect. In terms of destination, relative to Bogotá, only the Island region has a positive and significant effect on the conditional median of log-expenditure. Having children reduces the conditional median of expenditure, as expected given the time constrain that this variables imposes, while being male is associated with a positive effect.

The estimates of conditional quantile regressions for quantile levels at 0.1, 0.25, 0.75, and 0.90 reveal marked heterogeneity in the marginal effects of explanatory variables across the tourist expenditure distribution. A prominent example is the heterogeneity in income elasticity across different values of τ. Income is significant only for quantiles above the median, with a positive estimated marginal effect. The highest estimate is observed at $\tau = 0.75$, with a magnitude similar to that obtained in the mean regression. To provide insight into this finding, we can consider spenders in the left part of the conditional distribution as having more income restrictions. Consequently, a marginal change in their income is expected to be directed toward other types of goods and services rather than tourism spending.

Individual characteristics like age or having children appear to be more influential for quantiles below the median. In contrast, education exhibits positive and significant effects for quantiles above the median with the most pronounced estimates observed at $\tau = 0.75$ and 0.90. This heterogeneous effect of education aligns with the perspectives of Hung et al. (2012) and can be attributed to the differential information search costs and the accumulation of cultural capital experienced by heavy spenders. Across all quantile levels studied, being male is associated

[5]The methodology requires the explanatory variables in the outcome equation to be a subset of the explanatory variables in the selection equation. To include destination and motive for non-travelers, we generated random values to complete the missing values for non-travellers. The results in both equations remained robust to this manipulation.

with higher expenditure. In terms of motives and destination, shopping tourists have higher expenditures in all quantiles compared to business tourists. However, tourists traveling to regions other than Bogotá tend to have lower expenditures, with the exception of the Island region, which exhibits higher expenditure, but only for quantiles below the median.

In general, the results from the conditional quantile regressions confirm the presence of heterogeneous effects depending on the level of tourist expenditure. Among heavy spenders, income and education emerge as significant factors determining spending patterns. In contrast, individuals in the lower part of the conditional distributions appear to be influenced more by sociodemographic variables such as age or having children, which can be related to preferences and time restrictions. Gender, motivation, and destination exhibit significant effects across all quantiles.

6 Conclusions and future research

The analysis presented in this paper identified the main determinants of both travel decisions and tourist spending in Colombia, leveraging individual-level data. We explore the heterogeneity of the marginal effects of a set of explanatory variables across various conditional quantiles of tourist expenditure while accounting for the detected selection bias. Our findings are aligned with previous research in other contexts and confirm the high complexity of the tourism service. Based on the results of the selection equation, it is evident that domestic tourism is not pursued by a significant portion of the Colombian population and that factors such as income, time constraints, information availability, and personal preferences contribute to determining the decision to become a tourist. Moreover, these factors exert varying degrees of influence depending on the level of tourist expenditure. Among heavy spenders, income and education emerge are relevant determinants; in contrast, lower quantiles of the distribution are more affected by age or the presence of children. Trip-related factors such as motivation and destination were also found to be relevant.

Our findings offer policy-relevant insights for both destination managers and policymakers interested in promoting travel decisions and increasing economic revenues from domestic tourist activities. To achieve these objectives, some initiatives are feasible. For instance, it is necessary to reduce the information-search costs, particularly for individuals with lower levels of education, through the implementation of properly designed marketing strategies. This can enhance their access to travel-related information and potentially encourage them to become tourists. Other alternative is to develop tailored tourist packages and amenities for individuals facing income and time constraints. These initiatives should be accompanied

Table 3: Conditional quantile regression estimates correcting for sample selection

VARIABLES	Arellano-Bonhomme estimates					
	Selection	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
A. Sociodemographic						
Log-income	0.0402***	-0.0149	-0.0061	0.0221*	0.0327**	0.0206*
Age	0.0072**	0.0126**	0.0048*	0.0009	0.0109	0.0312
Age^2	-0.0001***	-0.0001*	0.0000	0.0001	-0.0000	-0.0002
Male	0.0392**	0.1581***	0.0867***	0.1574**	0.3390**	0.3674**
Strata 1	0.2235	0.3011	-0.6773	0.1704	0.1623	-0.1634
Strata 2	0.3036	0.4520	-0.4824	0.4376	0.1018	-0.2756
Strata 3	0.4801**	0.4519	-0.4824	0.3931	-0.0210	-0.3219
Strata 4	0.6444***	0.5226	-0.5000	0.4295	-0.1315	-0.6207
Strata 5	0.6728***	0.6638	-0.3172	0.5777	0.1657	-0.4578
Strata 6	0.7504***	0.6669	-0.3172	0.5198	0.1658	-0.3544
Children	-0.0798***	-0.0742**	-0.0688***	-0.0115**	-0.0447	-0.0293
Recreational house	0.6431***	-0.6082	-0.6728	-0.6449	-0.6973	-0.7714
B. Education (Ref: None)						
Primary	0.1043*	0.1445	0.3110	0.2587	1.3418	1.3558**
Secondary	0.1241*	0.1854	0.4906	0.4062	1.5026	2.9115**
Upper secondary	0.2131***	0.2240	0.4646	0.4047	1.3638**	1.9808***
Superior	0.4963***	0.2290	0.4412	0.3511**	1.4928***	1.8984***
C. Motive (Ref: Business)						
Vacation		-0.1303**	-0.2052***	-0.2376***	-0.1611*	-0.2961
Familiar		-0.6744***	-0.7406***	-0.6860***	-0.7149***	-0.7822***
Education		0.2094	0.2019	-0.1954	-0.4716	-0.1221
Health		-0.1796	-0.3944*	-0.4149*	-0.4887*	-0.4994
Religion		-0.5035	-0.4200	-0.6497*	-1.0248**	-1.4928**
Shopping		0.7546***	0.9295**	1.4038**	0.8224*	0.2955*
Other		-0.0634	-0.3225	-0.1728	-0.7534	-0.7734
D. Destination (Ref: Bogotá)						
Coffee region		-0.5226***	-0.4435***	-0.5785***	-0.6948***	-0.7024**
Caribbean region		-0.2399	-0.1188*	-0.2511*	-0.3091*	-0.3969
Island region		1.1878	0.9476**	0.5337***	-0.2263	-0.1341
Pacific region		-0.4638***	-0.4512***	-0.6061***	-0.8479***	-0.9714***
Antioquia		-0.0163	-0.0268	-0.0821	-0.3661	-0.7223
Eastern region		-0.3975**	-0.4211**	-0.4917***	-0.8084***	-1.1618**
Other		-0.4541***	-0.3998***	-0.60811***	-0.6925***	-1.1920***
E. Exclusion restriction						
Quarter 2	-0.0415**					
Quarter 3	-0.1034***					
Quarter 4	-0.1531***					
F. Dependence parameter						
ρ	0.7199**					

Notes: The table contains the results of the conditional quantile regressions correcting for sample selection. Individuals whose answer to strata and education questions was "not knowing" were excluded from the sample. Individuals with a daily per-person expenditure smaller than 0,30 U.S dollars were treated as non-travelers those. Individuals with kindergarden education and no education were included in the same group. In all the regressions we control for Region of residence, marital status and occupation. p-values for significance tests are represented as *** p<0.01, ** p<0.05, * p<0.1.

by stronger coordination between policymakers and destination managers, focusing on the provision of quality infrastructure and the diversification of tourist services. This approach should take into account the diverse preferences and profiles of tourists. By implementing these strategies, it is possible to stimulate domestic tourism and boost economic revenues in the sector.

Several avenues for future research are worth exploring. First, it is essential to apply empirical methodologies to different waves of the EGIT to analyze the dynamics of domestic tourist decisions. This becomes particularly interesting post-COVID-19, as it allows for the investigation of possible structural changes in tourist behavior. Our paper can serve as a foundational reference point for such research. Second, while our paper focused on aggregate spending, there is value in analyzing spending categories individually. This approach can provide a more comprehensive understanding of the phenomenon by introducing a new level of heterogeneity into the analysis. Finally, our current methods rely on the linearity of econometric models. However, non-linearities may play a significant role in this context, and further research in this direction could enhance our understanding of the complex relationships involving tourist behavior in Colombia.

References

Alzate, C. and Espinal, N. (2018). Estudio de las características del turismo doméstico en Colombia a partir de la matriz Origen-Destino. *Ensayos de Economía*, 28(52):151–164.

Arellano, M. and Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1):1–28.

Asgary, N., De Los Santos, G., Vincent, V., and Davila, V. (1997). The determinants of expenditures by mexican visitors to the border cities of texas. *Tourism Economics*, 3(4):319–328.

Azam, M. (2022). Microeconomic determinants of domestic tourism expenditure in india. *IZA Discussion Papers, Institute of Labor Economics (IZA), Bonn*, 15245.

Barquet, A., Brida, J. G., Osti, L., and Schubert, S. (2011). An analysis of tourists' expenditure on winter sports events through the tobit censorate model. *Tourism Economics*, 17(6):1197–1217.

Bel, F., Lacroix, A., Lyser, S., Rambonilaza, T., and Turpin, N. (2015). Domestic demand for tourism in rural areas: Insights from summer stays in three French regions. *Tourism Management*, 46:562–570.

Biewen, M. and Erhardt, P. (2021). arhomme: An implementation of the Arellano and Bonhomme (2017) estimator for quantile regression with selection correction. *The Stata Journal*, 21(3):602–625.

Borah, B. and Basu, A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Econ*, 22(9):1052–1070.

Borjas, G. (1980). The relationship between wages and weekly hours of work: The role of division bias. *The Journal of Human Resources*, 15(3):409–423.

Brida, G., Bibiana, L., Moreno, L., and Santiñaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay. *Tourism Management*, 69:62–68.

Brida, G., Cortés-Jiménez, I., and Pulina, M. (2016). Has the tourism-led growth hypothesis been validated? A literature review. *Current Issues in Tourism*, 19(5):394–430.

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

Brida, J. G., Fasone, V., Scuderi, R., and Zapata-Aguirre, S. (2014). Research note: Exploring the determinants of cruise passengers' expenditure at ports of call in uruguay. *Tourism Economics*, 20(5):1133–1143.

Cannon, T. F. and Ford, J. (2002). Relationship of demographic and trip characteristics to visitor spending: an analysis of sports travel visitors across time. *Tourism economics*, 8(3):263–271.

Figini, P. and Vici, L. (2010). Tourism and growth in a cross section of countries. *Tourism Economics*, 16(4):789–805.

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

Góngora, A. and Osorio, A. (2020). Tu edad sí importa: Factores sociodemográficos que inciden en la elección de viajar por turismo doméstico en Colombia. *Thesis, Universidad Santo Tomás*.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. (2010). *Selection Bias and Self-Selection*. The New Palgrave Economics Collection, Palgrave Macmillan, London.

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).

Jang, S. S. and Ham, S. (2009). A double-hurdle analysis of travel expenditure: Baby boomer seniors versus older seniors. *Tourism management*, 30(3):372–380.

Kim, W. G., Kim, T., Gazzoli, G., Park, Y., Kim, S. H., and Park, S. S. (2011). Factors affecting the travel expenditure of visitors to macau, china. *Tourism Economics*, 17(4):857–883.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Lew, A. A. and Ng, P. T. (2012). Using quantile regression to understand visitor spending. *Journal of Travel Research*, 51(3):278–288.

Mak, J., Moncur, J., and Yonamine, D. (1977). Determinants of visitor expenditures and visitor lengths of stay: A cross-section analysis of us visitors to hawaii. *Journal of Travel Research*, 15(3):5–8.

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

Mitra, J., Wilson, C., Managi, S., Kler, P., Prayaga, P., and Khanal, U. (2019). What determines whale watching tourists' expenditure? A study fro Hervey Bay, Australia. *Tourism Economics*, 25(7):1134–1141.

Park, S., Woo, M., and Nicolau, J. (2020). Determinant factors of tourist expenses. *Journal of Travel Research*, 59(2):267–280.

Pérez-Rodríguez, J. V. and Ledesma-Rodriguez, F. (2021). Unconditional quantile regression and tourism expenditure: The case of the canary islands. *Tourism Economics*, 27(4):626–648.

Sahoo, B. K., Nayak, R., and Mahalik, M. K. (2022). Factors affecting domestic tourism spending in india. *Annals of Tourism Research Empirical Insights*, 3(2):100050.

Seiler, V. L., Hsieh, S., Seiler, M. J., and Hsieh, C. (2003). Modeling travel expenditures for taiwanese tourism. *Journal of Travel & Tourism Marketing*, 13(4):47–60.

Sharma, A., Woodward, R., and Grillini, S. (2020). Unconditional quantile regression analysis of UK inbound tourist expenditures. *Economics Letters*, 186.

Song, H., Dwyer, L., and ZhengCao, L. (2012). Tourism Economics Research: A Review and Assessment. *Annals of Tourism Research*, 39(3):1653–1682.

Thrane, C. (2016). The determinants of Norwegian's summer tourism expenditure: foreign and domestic trips. *Tourism Economics*, 22(1):31–46.

Thrane, C. and Farstad, E. (2011). Domestic tourism expenditures: The non-linear effects of lenght of stay and travel party size. *Tourism Management*, 32:46–52.

Wang, Y. and Davidson, M. C. (2010). A review of micro-analyses of tourist expenditure. *Current issues in Tourism*, 13(6):507–524.

Wang, Y., Rompf, P., Severt, D., and Peerapatdit, N. (2006). Examining and identifying the determinants of travel expenditure patterns. *International Journal of Tourism Research*, 8(5):333–346.

Zheng, B. and Zhang, Y. (2013). Household expenditures for leisure tourism in the usa, 1996 and 2006. *International Journal of Tourism Research*, 15(2):197–208.