Modeling tourism flows through gravity models: A quantile regression approach

Santeramo, Fabio Gaetano and Morelli, Mariangela

University of Foggia, Italy, Duke University, USA

1 April 2015

Modeling tourism flows through gravity models:

A quantile regression approach

Santeramo Fabio Gaetano
University of Foggia

Morelli Mariangela
Duke University

Abstract

Gravity models are widely used to study tourism flows. The peculiarities of the segmented international demand for agritourism in Italy is examined by means of novel approach: a panel data quantile regression. We characterize the international demand for Italian agritourism with a large dataset, by considering data of thirty-three countries of origin, from 1998 to 2010. Distance and income are major determinants, but we also found that mutual agreements and high urbanization rates in countries of origin are associated with larger flows of incoming tourists.

KEYWORDS: *Flows, Arrivals, Agritourism, Gravity, PPML, Quantile Regression*

JEL : L83, Q17, F22

Corresponding author:
Fabio Gaetano Santeramo
University of Foggia
Via Napoli 25, 71100, Foggia, Italy
Email: fabiogaetano.santeramo@gmail.com

A revised version of the present paper has been accepted for publication in Current Issues in Tourism. Suggested citation: Santeramo, F.G., Morelli, M. (2015) Modeling tourism flows through gravity models: A quantile regression approach. Current Issues in Tourism. Doi: dx.doi.org/10.1080/13683500.2015.1051518

Modeling tourism flows through gravity models: A quantile regression approach

Introduction

The changing scenario of the tourism sector has both encouraged the demand for agritourism at national and international level (Santeramo, 2015), and catalyzed the expansion of the supply of agritourism (Barbieri & Mshenga, 2008). During the last decade, the number of foreign visitors in Italian agritourism has increased by two hundred percent and, more than one third of tourists hosted by Italian farmhouses are of foreign nationality (ISTAT, 2010). The international demand for Italian agritourism is very heterogeneous, although the main partners (Germany, United Kingdom, USA, Netherlands, France, Switzerland) account for eighty percent of the total international demand of agritourism in Italy (ISTAT, 2010).

Gravity models are commonly used to study tourism flows (Khadaroo & Seetanah, 2007; Fourie & Santana-Gallego, 2011; Massidda, Etzo, & Piras, 2015), although recent advances in econometric modeling have broadened the set of results that can be derived through gravity-type analysis (Hall, 2012). We propose to estimate quantile regression (Koenker, 2005) to study demand segmentation.

Following a preliminary study conducted by Santeramo (2015) on the international demand for Italian agritourism, we provide numerous further insights. We disentangle the determinants for incoming tourists to visit Italy and, in particular, to choose Italian agritourism. Moreover, we take into account both arrivals and durations of stay in order to achieve a broader picture of the phenomenon. Finally, we explore how Italian agritourism may satisfy different segments of the international demand.

The empirical analysis

According to the gravity model, the bilateral volume of flows among countries is proportional to the "mass" of the countries (e.g. Gross Domestic Product per capita), and inversely related to their respective distance:

(1) $\ln X_{ij} = \alpha_0 + \alpha \ln Y_i + \beta \ln Y_j - \gamma \ln D_{ij} + \varepsilon_{ij}$

where G is a scale factor, X_{ij} represents the trade or migration flow, Y_i and Y_j proxy the economic masses of country of origin (i) and country of destination (j), and D_{ij} is the distance between the two countries. We follow Silva and Tenreyro (2006) to account for zero flows and heteroskedasticity in the error term; hence we adopt a pseudo-Poisson Maximum Likelihood estimator (PPML), with the following set of first-order conditions:

(2) $\sum_{i=1}^{k}(X_k - \exp(Z_k\hat{\alpha})) = 0$

where $X_k$represents trade flows, Z_k is the full vector of explanatory covariates, $\exp(Z_k\hat{\alpha})$ is the expected value of $X_k$conditional on covariates (i.e. $E[X_k|Z_k]$). Wooldridge (2002, p. 676) argues that PPML Z_k is consistent if the conditional mean is correctly specified, that is if $E[X_k|Z_k] = \exp(Z_k\hat{\alpha})$ holds. The property applies regardless of the count data adopted. We also introduce a relatively novel approach by estimating gravity-type models through quantile counts regression techniques (Machado & Silva, 2005). Standard linear regression techniques synthesize the average relationship among dependent and independent variables, based on the conditional mean function$E[X_k|Z_k]$. Such a relationship is able to provide a narrow view: we might be interested in describing the relationship at different points in the conditional distribution X_k.For instance, the median function $Q_q[X_k|Z_k]$ describes the relationships at the median point (or 50th percentile, or quantile Q2), of the empirical distribution. Moreover, the quantile regression can be used to model conditional quantiles of the joint distribution of X_k and Z_k at selected quantiles. A further advantage is

that quantile regression is more robust to outliers than least squares regression, and does not rely on assumptions of the parametric distribution of the error process. Finally, the quantile regression estimator is asymptotically normally distributed (Koenker, 2005). Finally we test for statistical differences in quantile coefficients: p-values smaller than 0.1 indicate that the relationship is different for small, median, and large tourist flows. Intuitively this implies that policy makers should pay attention in differentiating marketing campaigns based both on the intrinsic characteristics of the country and on the observed tourist flows[a]. To the best of our knowledge, this is the first empirical application of quantile regression in tourism economics.

The model has been applied to a data-set on touristic flows from 1998 to 2010 for thirty-three countries of origin: we account for more than ninety percent of the total agritourism flows to Italy. The dependent variable is the number of arrivals of foreigners to Italian agritouristic structure. The data was extracted from the database of the Italian Institute of Statistics (ISTAT). The total number of agritouristic structures and the number of beds are accurate proxies of the supply. The gross domestic product at purchasing power parity per capita, expressed in current U.S. dollars, was extracted from the World Economic Outlook Database of International Monetary Fund. Data for population, in millions of habitants, was obtained from the FAO database. The geographical distance among capitals, expressed in kilometers, is computed using the Haversine formula and coordinates extracted from the CIA's The World Factbook. Descriptive statistics are reported in table 1.

< TABLE 1 ABOUT HERE >

As total GDP may overestimate the effect of the Italian supply for tourism, we have proxied it with the number of touristic and agritouristic structures[b]. We expect a positive relationship with the number of arrivals and the total duration of stay.

On the demand side, the countries of origin's purchasing capacity have been proxied by per capita GDP at PPP, while the effect of the economy size is captured by the total population. We expect a positive sign for both determinants. The expected signs of the variables "Rurality" and "Agricultual-Pop" may be ambiguous. In line with Santeramo (2015), we expect negative relationships: the higher the percentage of population living in rural areas, and working in agricultural sector, the lower the demand for agritourism.

The geographical distance between Italy and the country of origin proxies travel costs. Although distance is the main friction to international flows (Disdier & Head, 2008), transaction costs may play a significant role: dummies on international agreements are adopted to proxy those costs. Sharing the same currency and have Schengen agreement should facilitate movements of tourists. However if and how effective are these frictions for the international demand of agritourim is an empirical question.

Results

The results on the factors influencing the international demand for Italian agritourism are reported in Table 2. We present the final specification on four different dependent variables: number of arrivals and duration of stay in the general touristic sector, number of arrivals and duration of stay in agritourism.

< TABLE 2 ABOUT HERE >

Results show that the Italian supply is a major determinant for rural tourism (the variable is statistically significant at 1% level), but this is not generally true for tourism. Intuitively, being a small share of Italy's touristic sector, demand for agritourism can be incremented by expanding the supply. Therefore, Italy should increase the proportion of its agritourism resources *vis-a-vis* the usual touristic structures in order to increase the international

demand for tourism in Italy. The results are in line with those presented in Khadaroo and Seetanah (2007) who found that tourist arrivals change less than proportionally with change in number of infrastructures.

As for the demand side, all variables are statistically significant (except for "Per-capita GDP" in touristic sector), but not all determinants have the same importance. We found that the richer the Countries, and the richer the tourists, the higher the demand for Italian agritourism. This is not clear cut for the whole touristic sector. Moreover, the larger the population, the higher the demand for tourism in Italy will be. Our study suggests Italian entrepreneurs should expand their business by targeting populous Countries, possibly with a solid growth in income per-capita. The variables "Rurality" and "Agricultural-Pop" show that the higher the percentage of population living in rural areas (say 1% increase), and working in the agricultural sector, the lower the demand for tourism. The effect is twice as large for rural tourism[c].

As for frictions, we found that "Distance" is negative and statistically significant at 1% level. Distance affects rural demand more than general demand for tourism, but is equally influential on the duration of stay. While the variable "Euro" is not statistically significant, the "Schengen agreement" has enhanced international demand for tourism.

< TABLE 3 ABOUT HERE >

Table 3 summarizes the estimation by mean of quantile regressions. We present the results obtained at three quantiles: the 25^{th} percentile, the 50^{th} percentile and the 75^{th} percentile. The relationships among number of structures and agritourism and distance are stronger at the median point than at different quantiles. This implies that the level of income and the travel costs are not major determinants in explaining small and large flows.

Table 3 also reports the results of the statistical tests for difference in quantile

coefficients. P-values smaller that 0.1 indicate that the variable has (statistically significant) different influences on tourist flows. We found that the effect of distance is different for small and large flows, and the latter are less affected. The results are confirmed for the entire touristic flows as well as for agritourism. Two other results are particularly interesting. First, the expansion of agritourism supply would mainly attract visitors from Countries that are currently registering limited (and median) outgoing flows to Italy. Second, we found that the effects of "Rurality" and "Agricultural-Pop" are statistically different across quantiles: the sum of significant coefficients for the two variables at the first and third quantile is respectively equal to -0.18 and -0.25 (with -0.08 at median value). This implies that small and large incoming flows to Italian agritourism resources are substantially correlated with the degree of rurality of the country of origin, whereas this dynamic is not observed for "average" incoming flows.

Conclusions

Our empirical investigation presents a novel approach to analyze tourism flows. We have explored the international demand for Italian agritourism by estimating gravity-type models through quantile regression. By testing for differences in the quantile coefficients we show how to characterize the international demand: these findings cannot be captured by standard techniques (e.g. OLS or PPML estimation) as they provide average effects.

We found that, while the touristic sector is mature in Italy, agritourism is still a supply-driven niche. Agritourism demand is very reactive to the GDP per capita and the Population of countries of origin. In line with Khadaroo and Seetanah (2007) we found that distance is a main friction to tourist flows, but its effect is different for small and large touristic flows: the latter are less affected. Reduced cultural and economic distances due to shared political

agreements (e.g. Schengen agreements, or same currency) tend to facilitate tourism flows. This result is attested also in Yang and Wong (2012). We also found that the larger the urbanization the higher the probability of attracting new tourists would be.

Rural tourism industry and conventional tourism differ: while the supply of agritourism resources should be expanded, the same is not advisable for the other tourist companies; visitors from developed and highly urbanized countries are very likely to prefer agritourism, and should be targeted by ad hoc marketing campaigns[d].

The analysis is not exempt from limitations. First, as recently pointed by Guizzardi and Bernini (2012), underreporting is an important issue in official sources on tourism. However, while underreporting generally weakens the significance level of our estimates, the direction of causality is unaltered, and the implications we provide are still valid. Second, the use of the gravity model itself suffers from two main limitations: the gravity model is unable to take into account the effects of changes in travel costs in that they are proxied by a time-invariant variable (distance); in addition, the gravity model fails to capture out-of-average dynamics. The latter limitation has been partially overcome by mean of the quantile regression analysis, but the former limitation still applies. We add novelty to the current knowledge in tourism economics, leaving open the debate to future research.

References

Barbieri C., & Mshenga P. M. (2008). The role of the firm and owner characteristics on the performance of agritourism farms. *Sociologia Ruralis*, 48(2), 166-183.

Brown, D. (2005). Rural Tourism: An Annotated Bibliography. *Technical Report, Economic Research Service, U.S. Dept. of Agriculture*.

Butler, R., C. Hall, & J. Jenkins. (1998). *Tourism and Recreation in Rural Areas*. New York: Wiley.

Disdier, A. C., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics*, 90(1), 37-48.

Fourie, J., & Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management*, 32(6), 1364-1370.

Guizzardi, A., & Bernini, C. (2012). Measuring underreporting in accommodation statistics: evidence from Italy. *Current Issues in Tourism*, 15(6), 597-602.

Hall, C.M. 2012, Spatial analysis: A critical tool for tourism geographies, pp.163-173 in Wilson (ed.). *The Routledge Handbook of Tourism Geographies*. London: Routledge,.

Istat (2010) Movimento dei clienti nelle strutture ricettive.

Khadaroo, J., & Seetanah, B. (2007). Transport infrastructure and tourism development. *Annals of Tourism Research*, 34(4), 1021-1032.

Koenker, R. (2005). Quantile regression (No. 38). Cambridge university press.

Kožić, I. (2014). Detecting international tourism demand growth cycles. *Current Issues in Tourism*, 17(5), 397-403.

Machado, J. A. F., & Silva, J. S. (2005). Quantiles for counts. *Journal of the American Statistical Association*, 100(472), 1226-1237.

Massidda, C., Etzo, I., & Piras, R. (2015 forthcoming). Migration and inbound tourism: an Italian perspective. *Current Issues in Tourism*, 1-20.

Phillip, S., Hunter, C., & Blackstock, K. (2010). A typology for defining agritourism. *Tourism Management*, 31(6), 754-758.

Santeramo, F. G. (2015). Promoting the international demand for agritourism: empirical evidence from dynamic panel. *Tourism Economics*, doi: dx.doi.org/10.5367/te.2014.0397

Salman, A.K. (2003) Estimating Tourist Demand through Cointegration Analysis: Swedish Data. *Current Issues in Tourisms*, 6(4), 323-339.

Silva, S., & Tenreyro S. (2006). The Log of Gravity. *The Review of Economics and Statistics,* 88(4), 641-658.

Tchetchik, A., Fleischer, A., & Finkelshtain, I. (2008). Differentiation and synergies in rural tourism: estimation and simulation of the Israeli market. *American journal of agricultural economics*, 90(2), 553-570.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yang, Y., & Wong, K. K. (2012). The influence of cultural distance on China inbound tourism flows: a panel data gravity model approach. *Asian Geographer*, 29(1), 21-37.

Table 1 – Definition of variables and descriptive statistics

Variable name		Mean	Std.dev
$Arrivals_{it}$	Foreign tourists in Italian structures from country j in year t (in .000 absolute value).	187.6	44.1
$Arrivals\_AGR_{it}$	Foreign tourist in Italian agritourisms from country j in year t (in .000 absolute value).	13.4	37.9
Per-capita GPD_{it}	GDP per capita (current U.S. dollars) of country j, year t	22.4	16.4
$Population_{it}$	Total population (in millions) of country j, year t	99.1	262.3
$Distance_{ij}$	The distance between Italy and country j in .000 kilometers	4.2	4.5
$Structures_{it}$	Number of Italian touristic structures (.000) in year t	112.7	29.1
$Agritourisms_{it}$	Number of Italian agrituristic structures (.000) in year t	9.8	4.2
$Euro_{i}$	1if country j has adopted the euro, 0 otherwise	0.30	0.46
Schengen $agreement_{i}$	1if country j has signed Schengen agreement, 0 otherwise	0.16	0.37
$Agricultural\text{-}Pop_{it}$	Agricultural population (in percent) of country j, year t	3.6	7.6
$Rurality_{it}$	Agricultural population (in percent) of country j, year t	26.5	13.6

The statistics are computed from a pooled sample.

Table 2 –Determinants of foreign arrivals in Italian structures

	Touristic sector		*Agritouristic sector*	
Dependent variable:	Arrivals	Duration of stay	Arrivals	Duration of stay
Supply				
Number of structures	0.027	0.015		
	(1.34)	(0.86)		
Number of agritourism			0.139	0.090
			(6.49)**	(5.05)**
Demand				
Per-capita GPD	0.001	0.000	0.016	0.011
	(0.19)	(0.04)	(2.50)*	(2.06)*
Population	0.062	0.056	0.077	0.063
	(13.07)**	(13.07)**	(11.15)**	(10.40)**
Rurality	-0.002	-0.003	-0.005	-0.004
	(6.02)**	(6.73)**	(9.27)**	(9.00)**
Agriculture*Rurality	-0.011	-0.008	-0.017	-0.013
	(7.17)**	(6.12)**	(7.06)**	(6.62)**
Frictions				
Distance	-0.102	-0.112	-0.130	-0.116
	(13.28)**	(16.88)**	(11.03)**	(11.28)**
Schengen agreement	0.036	0.030	0.073	0.058
	(2.92)**	(2.57)*	(3.96)**	(3.76)**
Euro	0.019	0.011	0.025	0.025
	(1.50)	(0.84)	(1.39)	(1.58)
Constant	2.294	2.719	0.963	1.671
	(9.62)**	(13.15)**	(4.34)**	(8.96)**
R^2	0.59	0.63	0.59	0.57
Observations	466	466	466	466

+ $p<0.1$; * $p<0.05$; ** $p<0.01$

Table 3 – Quantile regression of determinants on arrivals of foreign visitors in Italian structures

			Touristic sector				*Agritouristic sector*	
Dependent variable:	Q1	Q2	Q3	Test Q1=Q2=Q3	Q1	Q2	Q3	Test Q1=Q2=Q3
Supply								
Number of structures	0.044	0.051	0.030	P-value: 0.638				
	(2.05)*	(2.13)*	(1.01)					
Number of agritourism					0.158	0.192	0.116	P-value: 0.062
					(5.41)**	(6.06)**	(4.87)**	
Demand								
Per-capita GPD	0.001	-0.000	0.009	P-value: 0.496	0.023	0.006	0.032	P-value: 0.564
	(0.53)	(0.03)	(2.17)*		(4.25)**	(0.68)	(5.96)**	
Population	0.067	0.076	0.054	P-value: 0.128	0.070	0.081	0.083	P-value: 0.261
	(8.26)**	(13.02)**	(5.70)**		(5.36)**	(6.87)**	(12.30)**	
Rurality	-0.003	-0.002	-0.004	P-value: 0.021	-0.006	-0.008	-0.004	P-value: 0.097
	(6.26)**	(4.70)**	(5.35)**		(8.67)**	(7.29)**	(3.99)**	
Agriculture*Rurality	-0.009	-0.013	-0.006	P-value: 0.022	-0.012	-0.005	-0.021	P-value: 0.001
	(4.18)**	(4.28)**	(2.38)*		(4.04)**	(1.15)	(7.16)**	
Frictions								
Distance	-0.125	-0.133	-0.093	P-value: 0.028	-0.140	-0.182	-0.100	P-value: 0.003
	(15.49)**	(15.52)**	(7.03)**		(6.16)**	(9.96)**	(7.77)**	
Schengen agreement	0.022	0.026	0.022	P-value: 0. 801	0.044	0.103	0.039	P-value: 0.389
	(1.28)	(1.54)	(1.41)		(1.53)	(2.84)**	(1.80)+	
Euro	-0.011	-0.005	0.053	P-value: 0.001	-0.018	-0.003	0.045	P-value: 0.150
	(0.69)	(0.30)	(3.59)**		(0.62)	(0.09)	(2.14)*	
Constant	2.231	2.106	2.277		0.907	0.876	0.798	
	(9.05)**	(6.59)**	(6.39)**		(2.98)**	(2.93)**	(3.56)**	
Observations	466	466	466		466	466	466	

+ $p<0.1$; * $p<0.05$; ** $p<0.01$

[a] We are very grateful to an anonymous referee for having provided suggestions that lead to this improvement of our analysis.

[b] We have also included the total number of beds in touristic and agritouristic structures. The results are not different, therefore we have considered only the number of structures for the present analysis.

[c] For instance, dwellers of low rural countries, such as those from Belgium and Luxembourg, Australia, Israel, Argentina, United Kingdom, Germany, Venezuela, New Zealand, Denmark, Brazil, Sweden, are more likely to be customers of Italian agritourism. *Ceteris paribus*, tourists from South Africa, Slovak Republic, Portugal, Slovenia and China (the most rural countries in our sample) are less likely to choose Italian agritourism for vacation.

[d] We are grateful to a referee for having asked to stress this difference.