Chi Fung Lam
The Chinese University of Hong Kong

Jian Ming Luo
City University of Macau

Museumpark Revisit: A Data Mining Approach in the Context of Hong Kong

It is important for tourism managers to understand the synergistic effects between museums in tourism management. This study raises is concerned with the "museumpark" and re-examines the synergistic effects among museums in Hong Kong. Utilising online photos from social media, over 2500 photos from more than 300 visitors were collected and analysed quantitatively. Through limited dependent variable regressions, the results show that once a visitor visits one museum in the "museumpark", the probability that this visitor will visit another museum nearby increase significantly. Based on the result of this study, museum researchers can further investigate the precise synergistic effects between museums.

Chi Fung Lam
Department of Finance
The Chinese University of Hong Kong
Hong Kong
Email: kelvinlam@cuhk.edu.hk

Jian Ming Luo
Faculty of International Tourism and Management
City University of Macau
Avenida Padre Tomás Pereira Taipa, Macau.
Phone: +853 8590 2535
Email: kenny.luo@connect.polyu.hk; kennyluo@cityu.mo

Jian Ming Luo is an Assistant Professor in the Faculty of International Tourism and Management at City University of Macau. His recent publications have appeared in the highly prestigious international academic journals.

Background

The concepts of "industrial district", "cluster" and "museumpark" are very similar. Becattini (1990) defined industrial district as "a socio-territorial entity with is characterized by the active presence of both a community of people and a population of firms in one naturally and historically bounded area" while Porter (2000) defined clusters as "geographic concentrations of interconnected firms, specialized suppliers, service providers, firms in related industries, and associated institutions in a particular field that compete but also co-operate".

Despite the lack of a formal definition of "museumpark", in 1996, Jansen-Verkebe and Rekom studied a "museumpark" or "museumquarter" in Rotterdam. The "Rotterdam Museumpark" included a museum for modern architecture, a museum of natural history, a local art museum, and a gallery for temporary art exhibitions, and a fine arts museum and all of them were within walking distance (Jansen-Verkebe & Rekom, 1996). The authors argued that physical clustering would generate spill-over effects, such as synergistic marketing strategies, sharing common resources, and providing convenience. Yet, there is a paucity of research which tests this hypothesis and the significance of spill-over due to the "museumpark" is still unknown.

This study provides a preliminary estimation of the effect of spill-over. The primary research question is centred on how likely visitors are to visit further museums after having visited one. While Hong Kong has not advertised a "museumpark", there are several museums located within walking distance of each other in Hong Kong, i.e., the Hong Kong Space Museum, the Hong Kong Museum of Art, Hong Kong Science Museum, Hong Kong Museum of History, the Hong Kong Heritage Discovery Center and the Hong Kong 3D museum. Although Jansen-Verkebe and Rekom (1996) did not provide a formal definition of a "museumpark", this area, spanning from Tsim Sha Tsui East to Tsim Sha Tsui, fits the criteria of "industrial district" defined by Becattini (1990) and "cluster" defined by Porter

(2000). This research does not attempt to provide any formal definition, but to direct research interest on formalizing the definition.

Empirical strategy and results

This study applied P-DBSCAN (Kisilevich, Mansmann, & Keim, 2010) clustering technique to the initially collected dataset. The initial data collection process focused on photos from social media platform Flickr tagged with the keyword "museum." The choice of this social media platform was based on convenience, as Flickr provides free access to many important data, such as photos, locations, and timestamps (Vu, Luo, Ye, Li, & Law, 2018). We set the bounding box for the photo function with parameter values as $la_{min} = 113.887603$, $lo_{min} = 22.215377$, $la_{max} = 114.360015$ and $l0_{max} = 22.51446$ to cover the entire Hong Kong geographical area.

The search covered a ten-year period from 2007 to 2016. All photos, whose tag field contained the provided keyword and taken time within the specified time frame, were included. The search returned 2,843 photos about the museum from 349 visitors. The museums included were the Space Museum, Museum of Art, Museum of History, Science Museum, and Heritage Museum as other museums did not yield sufficient data. After examining the locations, the clusters were marked using the names of the corresponding museums. Since the location between the Museum of Art and the Museum of History is relatively close, the location cannot be distinguished from the data. Therefore, we considered them as one attraction.

All of the museums were located in Tsim Sha Tsui and Tsim Sha Tsui East area. Since some visitors can visit multiple museums at the same time, the total number of visitors can exceed 349. There were 321 visitors who visited one museum, 25 visitors who visited

two museums, and three visitors who visited three museums. Table 1 provides a summary of visit frequency for one and two museums.

Table 1: Frequency of Museum Visits

	Space	Art	History & Science	**No. of Visitors**
Visit 1 Museum	59	141	121	321
Visit 2 Museums	17	19	14	50

To analyze how prior visit affects the likelihood or probability of visiting other museums around the area, this study employed both a logit and a probit regression model. Since the dependent variable was a limited dependent variable, multiple linear regression would be subject to an out of boundary problem, i.e., the predicted value would be outside the 0 and 1 range, and heteroskedasticity problem (Kennedy, 2003). Despite the beautiful property of Ordinary Least Square (OLS) that it was robust to violations of some of the assumptions, OLS was not robust to the limited dependent variables unless the mean of the dependent variable was close to 0.5 (Walsh, 1987). Clearly, according to Table 1, the data in this study did not exhibit this property. The more appropriate way to study the relationship was to use a probit or logit model. A general linear model for limited dependent variable could be characterized

$$P(Y_i = 1|X_i) = H(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}) = H(X_i\beta)$$

where *H(.)* was the cumulative density function. When

$$H(z) = \frac{1}{1 + \exp(z)}$$

it was a logit model and when

$$H(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

it was a probit model. For a given set of data, the likelihood function was

$$L(\beta | X_i) = \prod_{i=1}^{n} H(X_i\beta)^{Y_i} (1 - H(X_i\beta))^{(1-Y_i)}$$

The dependent variables here were whether the visitor visited one of the museums, i.e., $Y_i=1$, if the visitor visited the corresponding museum and zero otherwise. Similarly, for the independent variables, $X_i = 1$, if the visitor visited the corresponding museum and zero otherwise. According to Kennedy (2003), as computational becomes relatively "cheap", both logit and probit model become the most common model to address latent or limited dependent variables. Since we did not have a prior hypothesis on how the visitation spread from one museum to another, we investigated three possibilities. Particularly,

$$Model\ 1: P(Space_i = 1|X_i) = H(\beta_0 + \beta_1 Arts_i + \beta_2 History_i)$$
$$Model\ 2: P(Art_i = 1|X_i) = H(\beta_0 + \beta_1 Space_i + \beta_2 History_i)$$
$$Model\ 3: P(History_i = 1|X_i) = H(\beta_0 + \beta_1 Space_i + \beta_2 Arts_i)$$

The parameters were estimated using maximum likelihood estimation with R. The result of the two models are presented in Table 2

Table 2 Logit and Probit Regression Results

		Model 1		Model 2		Model 3	
		Logit	Probit	Logit	Probit	Logit	Probit
Intercept	Coefficients	0.04753	0.0844	-0.8112	-0.47129	-0.6535	-0.3673
	std error	0.18313	0.11471	0.1538	0.09261	0.1578	0.0961
	p-value	0.795	0.462	1.34E-07	3.60E-07	3.46E-05	0.000132
Space	Coefficients			2.3148	1.28169	2.6468	1.4052
	std error			0.3423	0.19048	0.4076	0.2099
	p-value			1.36E-11	1.71E-11	8.40E-11	2.16E-11
Art	Coefficients	2.31478	1.21698			3.0158	1.666
	std error	0.34233	0.18064			0.3297	0.1696
	p-value	1.36E-11	1.62E-11			2.00E-16	2.00E-16
History	Coefficients	2.64677	1.35913	3.0158	1.69477		
	std error	0.40762	0.20079	0.3297	0.17175		
	p-value	8.40E-11	1.30E-11	2.00E-16	2.00E-16		
Log Likelihood		-156.477	-159.499	-198.425	-201.403	-183.097	-186.792
Chi-square		87.63	81.587	151.14	145.19	155.1	147.71

p-value	2.2e^-16	2.2e^-16	2.2e^-16	2.2e^-16	2.2e^-16	2.2e^-16

As Table 2 showed, all regressions, regardless of the underlying distribution, or the dependent variables, were significant at 1% significance and the signs of the coefficients, along with the Chi-square statistics, were consistent with the spill-over hypothesis that upon the visitation of one or more museums, the probability of the visitors visiting another museum increases.

Limitations and Conclusions

The methodology has limitations, for example, due to limited data, many important variables such as the cultural background, duration of stay of the visitors, etc. are not included in the model, which can cause the endogeneity problem. However, the most important finding is that upon visiting one or more museum, visitors tend to visit more museums. This result sharply contrasts with the results from Jansen-Verbeke and Rekom (1996), where the physical clustering of museums (the "museumpark" in Rotterdam) and the visiting of one museum dis did not increase the tendency to visit another museum. To conclude, this study calls attention to researchers and highlights the need for a more precise definition of "museumpark" and further examination of the spill-over effect among "museumpark" around the world.

References

Becattini, G. (1990). 4 The Marshallian industrial. Industrial districts, 37.

Jansen-Verbeke, M., & Rekom, J. (1996). Scanning Museum Visitor. Urban Tourism Marketing, Annals of Tourism Research, 23 (2), 364-375.

Kennedy, P. (2003). A Guide to Econometrics. Cambridge, MA: The MIT Press.

Kisilevich, S., Mansmann, F., & Keim, D. (2010). P-DBSCAN: a density based clustering algorithm for exploration and analysis of attractive areas using collections of geotagged photos. In Proceedings of the 1st international conference and exhibition on computing for geospatial research & application. Article no. 38, Bethesda, Maryland, USA.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. Economic development quarterly, 14 (1), 15-34.

Vu, H.Q., Luo, J.M., Ye, B.H., Li, G., & Law. R. (2018). Evaluating museum visitor experiences based on user-generated travel photos. Journal of Travel and Tourism Marketing, 35 (4), 493-506.

Walsh, A. (1987). Teaching, Understanding, and Interpretation of Logit Regression. Teaching Sociology, 15 (2), 178-183