



國立中山大學企業管理學系

Department of Information Management  
National Sun Yat-sen University

演唱會每票收益之變數分析  
—以支持向量機及兩階段最小平方法為例

An analysis of features of the average revenue of the concerts

林育慈

國立中山大學管理學院企業管理學系研究生

ritalinyutzu@g-mail.nsysu.edu.tw

指導教授：余健源 博士

cysheer@mail.nsysu.edu.tw

中華民國 108 年 3 月

March 2019

## 摘要

音樂演唱會的每票平均收益，向來是音樂產業內部進行票價訂定時的標準之一；假設某策展團體預計邀請一位從沒到過該國進行演出的音樂人時，預測其每票平均收益能夠幫助他們解決訂價問題，從其網路聲量、演出地點、是否發行新歌等變數出發，以機器學習之支持向量機將所有變數進行分類，以歸納並預測出該名音樂人之特徵別量是否屬於每票高平均收益族群。此外，本研究更將建立兩種不同模型來評估網路聲量是否能夠顯著解釋音樂演唱會之每票收益；第一種為：工具變數加入 Lasso Regression 中篩選後，再進行兩階段最小平方法之計量經濟學模型；第二種為單純地加入工具變數並直接進行兩階段最小平方法。藉此來探討將工具變數加入 Lasso Regression 中進行篩選的動作，是否能夠有效地提升計量模型之解釋度。

本研究透過蒐集兩年期之 Pollstar 資料，並建立在演唱會產業之業界觀點（覺醒音樂祭 Wake Up Festival / 傲頭厝展演空間 Our Town Love House），來發想此研究內容。研究結果顯示，據本研究所蒐集之變量特徵能以 99.406% 之預測能力來將音樂人族群進行分類。此外，並能發現在做兩階段最小平方法時，若是先將工具變數優先加入 Lasso Regression 來進行篩選，其所得之模型解釋力相較於第二種方法要來得高，證明 Lasso Regression 確實擁有篩選工具變數之能力。

**關鍵詞：**音樂產業、演唱會產業、機器學習、工具變數、套索算法、支持向量機、兩階段最小平方法。

## 壹、 研究方法

本研究採次級資料分析法，欲探究所蒐集之各種特徵變數對於演唱會每票平均收益之預測，並同時分析主要變數 Youtube 網路聲量是否對每票平均收益有顯著解釋力，且研究工具變數的加入是否能有有效幫助 Lasso Regression 進行變數篩選。資料分三部分蒐集，本研究以專門蒐集全球演唱會資料的 Pollstar 官方資料庫，從 2017 年 2 月至 2019 年 1 月演唱會之展演資料為樣本，作為音樂人在不同場次展演資訊之來源。音樂人名氣指標部分也以 Pollstar 所提供之每週釋出的百大收益排行榜為依據。此外，筆者亦手動以 Python 進行網路爬蟲來擷取 Youtube 瀏覽量資訊，並結合 Kword 網站所提供之 Youtube 追蹤資料來做使用。

## 貳、 研究結果

本研究使用 Pollstar 資料庫內的演唱會歷史資料作為主要研究變數，以及筆者網路爬蟲所得之 Youtube 資料與 Kword 資料，使用 R 語言之開放資源 e1071、 leaps、glmnet 與 Python 之開放資源來進行 Lasso Regression 機器學方法 SVM 支持向量機進行特徵分群預測。接著，再使用 R 語言之開放資源 AER、 leaps、glmnet、caret 來分別進行 Lasso Regression 及兩階段最小平方方法之研究，此階段分為兩種模型來探討，分別為「將工具變數加入 Lasso Regression」與「不將工具變數加入 Lasso Regression」。本章節將依序說明本研究之分析結果，包含敘述統計資料、支持向量機結果、第一種兩階段最小平方方法與第二種兩階段最小平方方法結果。

## 一、 資料敘述統計

本節筆者將針對本研究之特徵變數進行說明。資料期間自 2017 年 2 月 28 始，至 2019 年 1 月 31 日止，時間橫跨約一年半。由下表可發現演唱會平均會舉辦 1.18 場表演；平均售出票券為 17,793 張；平均最低票價為 411 美元，這可能是受到高價位搖滾區票券的影響；此外，演唱會之音樂人平均上榜 10.395 次，筆者推估是由於知名歌手經常舉辦巡迴演唱會所致。

接著，由表二可見，在本研究所蒐集之資料中，經 Live Nation 公司所策展之演唱會比例為 58.4%，該公司為全球演唱會市場之龍頭廠商；售票率部分，57.1%之演唱會皆能完售，推測是因筆者已排除舞台劇、馬戲團、低網路聲量音樂人以及低收益音樂人所致；65.6%的演唱會會邀請共演團體來進行暖場；85%的演唱會曾上過百大收益排行榜，此緣由筆者上述亦提到過。策展人部分，僅有 7.6%的演唱會未聘任策展公司來操作，僅靠表演場地之擁有者來進行演唱會推廣。另外，本研究所蒐集的資料 78.2%舉辦在美國，係因世界各地的音樂人皆會到美國進行巡迴演唱，且顯示出美國在音樂市場之強勢文化輸出；本資料僅有 5.1%舉辦於北美的加拿大、8.4%舉辦於歐洲、11.1%舉辦於美國南部、1.5%舉辦於南半球。接著，只有 20%的演唱會舉辦在美國連假，其餘皆辦於平日或普通的週休二日。值得關注的是，有 41.3%的演唱會之音樂人，再舉辦演唱會的近期內推出了新歌，這顯示發表新歌是音樂人舉辦演唱會的重要因素之一。

Table 1: Summary Statistic

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Shows	1.180	0.489	1	1	1	5
Month	7.359	2.320	2	5	9	12
Tickets_Sold	17,793	24,178	73	5,989	18,672	299,013
Capacity	15,152	14,041	180	6,164	17,162	80,413
Percentage	0.931	0.133	0.090	0.930	1.000	1.000
SoldOut	0.571	0.495	0	0	1	1
US_Gross	1,584,294	2,684,384	1	347,406	1,556,629	32,839,088
min_price	411	7,133	0.000	29.750	50.000	660,274
max_price	1,265	22,934	0	70.5	150	2,269,690
USGorss.TicketSold	8,517	26,623	0.018	65.824	9,362	478,774
Supporting_Dummy	0.656	0.475	0	0	1	1
Global_Dummy	0.850	0.357	0	1	1	1
Global.Times	10.395	12.627	0	1	22	45
InhousePromotion	0.076	0.265	0	0	0	1
no_promoter	0.721	0.722	0	0	1	5
US	0.782	0.413	0	1	1	1
Canada	0.051	0.221	0	0	0	1
Europe	0.084	0.278	0	0	0	1
UK	0.034	0.182	0	0	0	1
SouthUS	0.111	0.314	0	0	0	1
Oceania	0.015	0.121	0	0	0	1
Date_1_Whichday	10.190	8.337	1	4	16	31
Live.Nation	0.584	0.493	0	0	1	1
Holiday	0.200	0.400	0	0	0	1
Last	2.667	6.655	0	0	2	102
This	2.652	6.336	0	0	2	86
Change	0.058	0.444	-0.910	-0.120	0.100	4.930
new_song	0.413	0.492	0	0	1	1

Table 2: Summary of Dummy Variables

	no	yes
Live.Nation	0.416	0.584
SoldOut	0.429	0.571
Supporting_Act_Dummy	0.344	0.656
Global_Concert_Pulse_Dummy	0.150	0.850
InhousePromotion	0.924	0.076
US	0.218	0.782
Canada	0.949	0.051
Europe	0.916	0.084
UK	0.966	0.034
SouthUS	0.889	0.111
Oceania	0.985	0.015
Holiday	0.800	0.200
new_song	0.587	0.413

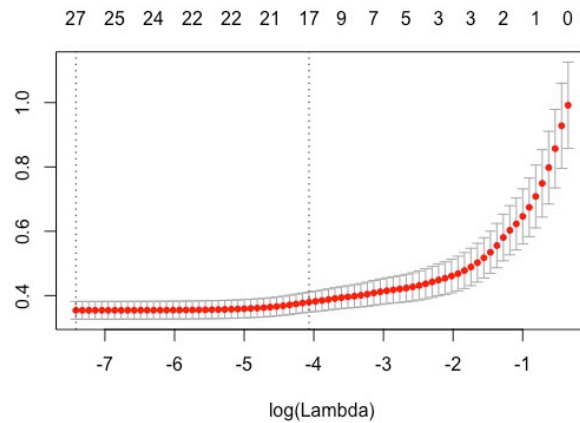
## 二、 分析結果：機器學習— SVM 支持向量機模型

進行 SVM 支持向量機前，筆者先使用 Lasso Regression 來處理所有變數，以擬合出最佳模型。此模型之懲罰 lambda 值為 0.0006003583，由表三可見演唱會上一週之 Youtube 網路聲量係數，被壓縮至 0。由圖一亦可觀察本筆研究資料之 lambda 壓縮情況。

接著，以機器學習中之 SVM 支持向量機來進行分群，以每票營收為應變數，並將所有變數進行分群，筆者主要觀察為所有變數在特徵化後是否能夠完整判斷出每票營收之不同族群的分類狀況。資料分類精確度為 99.40628%。

Table 3: Lasso Regression

Variables	Coefficient
(Intercept)	-864.79
Shows	-0.13
Year	0.21
Month	0.04
Tickets_Sold	1.41
Capacity	-0.39
Percentage	-0.03
SoldOut	-0.01
US_Gross	-0.35
min_price	0.01
max_price	-0.02
Supporting_Act_Dummy	0.00
Global_Concert_Pulse_Dummy	0.04
Global_Concert_Pulse_Times	-0.11
InhousePromotion	0.04
no_promoter	0.06
US	0.02
Canada	-0.02
Europe	0.00
UK	0.07
SouthUS	-0.02
Oceania	0.00
Date_1_Whichday	0.01
Live.Nation	-0.03
Holiday	-0.02
Last	.
This	0.00
Change	0.00
new_song	-0.03



圖（一）：Lasso Regression 之 lambda 值。

### 三、 進階分析結果：兩階段最小平方法結合工具變數（Model 1）與兩階段最小平方 法結合 Lasso Regression 與工具變數（Model 2）之比較

據表四（Model 1）可見，以計量經濟模型中的兩階段最小平方法結合工具變數來做迴歸，可發現上週網路聲量、六月、八月、九月、十月、十一月、十二月、售票率是否完售、以及總收益非常顯著。

據表五（Model 2）可見，進行兩階段最小平方法前，若先將工具變數同時加入 Lasso Regression 中先進行變數篩選，可發現上週網路聲量、六月、八月、九月、十月、十一月、十二月、售票率是否完售、以及總收益非常顯著。

然而，兩種模型之比較，據表六可發現，Model 2 之 Adjust R-squared 比 Model 1 來得更高，可證 Lasso Regression 在於結合工具變數進行變數篩選時，擁有較好的模型效果與解釋力。此外，筆者亦發現，Youtube 瀏覽量對於每票平均收益有高解釋力。

Table 4: Two-Stage Least Squares Model 1

		<i>Dependent variable:</i>	
		USGorss.TicketSold	
Last		2,360.632***	(790.606)
Year2018		12,840.760***	(718.928)
Month3		15,282.640**	(6,619.253)
Month4		15,123.880**	(6,416.835)
Month5		15,626.080**	(6,400.251)
Month6		17,310.020***	(6,403.528)
Month7		13,178.090**	(6,404.368)
Month8		16,728.940***	(6,402.798)
Month9		17,699.430***	(6,410.799)
Month10		17,811.560***	(6,406.890)
Month11		17,617.390***	(6,421.583)
Month12		20,012.920***	(6,456.607)
Tickets.Sold		0.863***	(0.009)
Percentage		412.614	(1,350.257)
SoldOut	4	-1,118.802***	(364.192)
log(US_Gross)		-2,602.278***	(156.696)
min_price		-17.182*	



Table 5: Two-Stage Least Squares Model 2

		<i>Dependent variable:</i>	
		USGorss.TicketSold	
Last		2,116.674***	(681.553)
Year2018		12,812.580***	(716.029)
Month3		15,468.900**	(6,580.017)
Month4		15,375.430**	(6,380.806)
Month5		15,880.530**	(6,363.216)
Month6		17,718.630***	(6,370.513)
Month7		13,593.170**	(6,368.706)
Month8		16,979.520***	(6,364.913)
Month9		17,951.320***	(6,373.096)
Month10		18,053.870***	(6,369.539)
Month11		18,306.470***	(6,395.262)
Month12		20,566.910***	(6,437.098)
Tickets_Sold		0.864***	(0.009)
Percentage		504.486	(1,345.878)
SoldOut	5	-1,259.703***	(362.469)
log(US_Gross)		-2,596.558***	(156.061)
min_price		0.080	

Table 6: Comparison of Model 1 and Model 2

Subject	Model.1	Model.2
Multiple R-squared	0.6063	0.6111
Adjusted R-squared	0.6046	0.6095
Wald Test	386.2	403.3
P-value	<2.2e-16	<2.2e-16

## 參、 結論與建議

結合 SVM 支持向量機，能夠高度預測音樂人之每票平均收益族群為何，共分為五個等級；此結果能夠為策展公司作為日後票券訂價之依據。此外，從計量模型角度來看，表演前的 Youtube 影片瀏覽量也確實能夠顯著地解釋每票平均營收。故筆者推論，若是策展團體在做利潤極大化與票券定價時，應參考其 Youtube 影片瀏覽量，且利用支持向量機來進行分群，可得到良好的商業模式。然而，礙於本研究之限制，Youtube 未提供各個音樂人之音樂表演類型，故無法將此特徵加入模型之中，若能加入此特徵變數，必定能夠再提高本研究之商業價值。

## 肆、 參考文獻

- A. Belloni, V. Chernozhukov, and C. Hansen (2010). Lasso methods for gaussian instrumental variables models. MIT Department of Economics Working Paper, No.11-14
- Mehmet Caner\* QingLiang(Michael) Fan (2010). The Adaptive Lasso Method for Instrumental Variable Selection. Department of Economics North Carolina State University.
- Akerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination\*. International Economic Review, 44(3), 1007-1040.
- Alan B. Krueger (2005). The Economics of Real Superstars: The Market for Rock Concerts in the Material World. Journal of Labor Economics, Vol. 23, No. 1, pp. 1-30.
- Alan B. Krueger & Marie Connolly (2005). Rockonomics: The Economics of popular music. National Bureau of economic research , JEL No. Z1, L82, O34.
- Caves, R. E. (2000). Creative industries: Contracts between art and commerce. Harvard University Press.
- Court, Andrew T. (1939). Hedonic Price Indexes with Automotive Examples. New York General Motors
- Dewenter (2012). On File Sharing With Indirect Network Effects Between Concert Ticket Sales and Music Recordings. Journal of Media Economics.