The effect of action movie production on the popularity of martial art styles——Using Multivariate Linear Regression

Junfei Sun sjf2002@sohu.com

June 2021

Abstract

This paper uses Multivariate Linear Regression and mathematics modeling to address the problem of how different dimensions of action movies featuring martial art styles would influence martial art spreading concerning the popularity of a specific martial art style and provides a new, unprecedented mathematical context to the influence of action movies to martial arts and therefore provides information on how detailed features in action movies are related to martial art popularity themselves. The study uses robust regression with significant exogeneity and without multicollinearity to eventually predict the result of the significant relationship of the historical background featured in that film and whether the design of the actions in that film is close to real combat are strongly correlated with the popularity of the martial art styles.

1 Introduction

In the film industry today, action movies play a large role and attract a large amount of attention from the audience. A large proportion of action movies involves combat scenes that feature specific martial art styles. Therefore, movie productions can consequently have an impact on the popularity of those martial art styles and eventually their spreading. For example, actors like Bruce Lee, Jackie Chan, and Jet Li have been considered to have produced movies that give Kung Fu three different kinds of cultural contexts and played significant roles in helping with the spread of Kung Fu (Wong, Suet-lan)(Farquhar, Mary, 2010).

There has been reasoning how would movie productions that feature martial art styles affect the martial art styles featured. However, most of them focus simply on whether those movie productions offer a positive or negative environment on martial art styles. As Dana White (President of UFC) has put"We owe big to gladiator and martial arts movies. Movies and video games paved the way for us. All we did was to bring fiction into reality." Lots of investigations show that movies have played helped the spreading of lots of martial arts by giving them a more idealistic context(Bolelli, Daniele, 2014). While some take movie's impact on martial art spreading negatively by stating that "extremely exaggerated interpretation and description of the movies may lead to the cognitive biases of the audience"(Li Yuan, 2013). However, what is missing from the previous study is specifically which part of those movie productions tend to be the ones making the impacts and which ones are relatively more insignificant.

Additionally, according to some interviews conducted with traditional Kung fu masters practicing different styles of Kung fu. There is a strong tendency of believing that action movies are playing a large role in promoting, or even saving traditional martial art and quoting: "Martial arts Movies makes a huge difference on the scale of people interested in martial arts and are the saviors of the awkward situation of traditional martial arts nowadays." This paper serves to verify such claims and discuss the influence of movie production towards martial arts spreading in detail by cutting the concept of movie production into small features.

Mathematically, Multivariate Linear Regression is used to verify the relationship and the significance between the explained variable we focus on (which is popularity in this paper) and the various features in movie productions of which we tend to verify the relationship with martial arts' popularity.

2 Notations

Most of the variables in this paper are not expressed using notations to make reading processes more convenient. Still, some notations needs to be clarified

Notations	Meaning
β_i	The coefficients for explanatory variables
λ	The coefficients for control variables
ϵ	random error term in regression model
VIF	Variance Inflation Factor
R^2	coefficient of determination
γ	coefficient for instrumental variables

Table 1: Notations table

3 Explained Variable

Popularity is set as the dependent variable as the variable explained. The popularity would be equal to the relative searching index of the name of the specific martial art and its synonyms in various languages on the largest searching Engine Google from 2004 Jan to 2021 May globally. The relative searching index (RSI) for different months would be modified and put into the following equation to generate the *Popularity*

$$Popularity = \frac{\sum_{i=1}^{209} RSI \times i}{209}$$

The equation above takes into consideration time on *Popularity* with the earlier data having a smaller weight over the 209 months. The weight increase linearly with time from $\frac{1}{209}$ to 1. This model is used to reflect how RSI nearer to the current time plays a large role in the popularity of a specific martial art style nowadays

The martial art styles we have selected as samples are the following: Kungfu, Karate, Boxing, Muay Thai, MMA, Silat, Jeet Kune Do, Bartitsu, Bokh, Angampora, Lethwei, Muay Lao Aikido, BJJ, Savate, Sambo, Kenpo, Arnis, Vale Tudo, dambe, judo, wrestling, Taekwondo, Hapkido, Kickboxing, Capoeira, Sumo, krav maga, Ninjutsu, and Kalaripayattu. It needs to be addressed that the styles above are not necessarily strictly spearated martial art styles but are conceptually separated. For example, there are lots of sub-styles with in Kung fu but Kung fu is still counted as one style as it is widely taken as a conceptual whole.

The Popularity value of them are shown below and the exact values refers to the Appendix

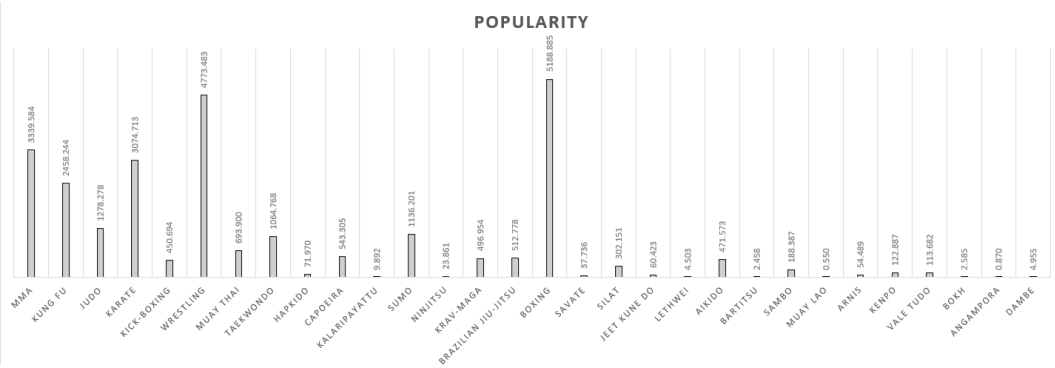


Figure 1: Popularity for different styles of martial arts

4 Determining explanatory and control variables

4.1 Explanatory Variables

Movie production in general is too broad of a concept. Therefore, some features that can be evaluated and quantified are extracted at the prerequisite that these features have the potential of being a factor influencing the popularity of a martial art style. Through thorough investigations on previous papers and interviews with martial artists that have experience in such fields, these features are extracted and quantified as explanatory variables: *Boxoffice*, *Movien*, *Combatvalue*, *Themetime*, *Nationality*,*TimeIndex* and *Mainstyle*. They represent different features of action movies featuring a specific martial art style. How do those variables serve the function of depicting the status of movie productions in details listed in the following table:

variable	variable description	values
Boxoffice (million)	average boxoffice for all the movies featuring a specific martial art style	R^+
Movien	number of movies featuring a specific martial art style (From 1930 till now)	N
Combatvalue	description of how close is the overall combat design to real combats for all movies featuring a specific martial art style	[0,1]
Nationality	evaluating whether the people displaying a specific martial art style in the movies are of the same nationality of the martial art style itself for all movies featuring that specific martial art style	[0,1]
Themetime (century)	average difference in centuries between the 21st century and the setting time for all movies featuring a specific martial art style	R^+
Mainstyle	evaluating whether a specific martial art style is a major style displayed in the movies featuring that style	[0,1]
TimeIndex	An index calculating the general time at which the movies of a specific martial art style are filmed, the larger the earlier the movies are made	$[0,+\infty)$

Table 2: table of explanatory variables

All the explanatory variables listed in table 2 are manually counted from movie databases: IMDB and Letterbox and the collected data are put in the appendix.

For *Combatvalue*, the combat value for each film is a dummy variable with the value of 0 or 1. 0 indicates the martial art style in the movie is far from real combats whereas 1 indicates the opposite. This approach is implemented as the combat value is hard to evaluate into specific numbers and easier to be categorized into binary dummy variables. Additionally, it is important to notice that there's no possibility of achieving 100% of similarity between the movie's action design and real combats gave that all action designs have been optimized and idealized for artistic values. Therefore, to determine whether an action design has combat value would equal to determining whether that action design is theoretically feasible and near real combats. Four features are the main factors deciding the combat value of action designs in this case: fluidity, wire-using, strength, and authenticity according to Master Zhang Ping (A Jeet Kune Do Shifu with experience in action design).

For instance, action design in Snake & Crane Arts of Shaolin (1978) would be categorized into 0 combat value given that it lacks fluidity even with a little amount of wire-using, strength, and authenticity. The actions in Crouching Tiger, Hidden Dragon(2000) would also be categorized into 0 combat value given that it lacks authenticity with the use of Chi and involves a large amount of

wire-using. On the other hand, Flash Point(2007) would be of 1 combat value as it fulfills all four main factors.

The overall *Combatvalue* for all the movies featuring a martial art style is the average of all the combat values for each movie, same for*Nationality* (which takes the average of dummy variables using 0 to indicate the nationality of the actor and the style doesn't match and 1 as the opposite) and *Mainstyle* (which takes the average of dummy variables using 0 to indicate the style not being a major style and 1 the opposite)

For *TimeIndex*, when are the movies majorly produced are thought to be a possible factor in the popularity of martial art as the influence of a movie on the audience would get varied through time. It's obtained through the following formula:

$$TimeIndex = \frac{\sum_{i=0} Movien_i \times i}{Movien_{total}}$$

in which i is the time gap in decades between the 2020s and the time a movie is filmed, $Movien_i$ showing the number of movies filmed in a decade with a time gap of i from 2020s. The specific movie distribution throughout time can be seen in the appendix.

Nationality is set as an explanatory variable for the fact that a specific kind of martial art is largely linked to the impression of its origin country, especially for those that cannot extract features from the martial art movements themselves to determine which style of martial art is presented. The nationality might be the most obvious signal for those audiences to link what they see to a specific martial art style(Ping Zhang, 2021).

Mainstyle is also considered to be relevant given that whether a martial art style is majorly featured will affect how much of that martial art style did the audience taken away after watching the movies.

All the explained and explanatory variables can be seen in the appendix.

4.2 Control Variables

After the explained variables are set, the control variables are determined to avoid endogeneity problems in the multiple regression. With the random error term containing all the variables that are linked to the dependent variables that are not listed out in the independent variables, The control variables are required to be both correlated with the random error term and the independent variables. Therefore, explained variables would avoid being endogenic and correlated with the random error term to a large extent. The control variables used in this regression are: *Olympics, Competition, Performance, and Game* The specific meaning and information on the control variables are listed in table 2 below

variable	variable description	Range
Olympics	evaluate whether a specific martial art style is a constant event in the Olmpics	0 or 1
Competition	evaluate the international sparring competitions of a specific martial art style in terms of internationality, history and number	R^+
Performance	evaluate the international performance competitions of a specific martial art style in terms of internationally, history and number	R^+
Game	describe how many major video games have featured a specific martial art style	N

Table 3: table of control variables

Game data are all from the IGDB database, Competition and Performance information come from the websites of corresponding events. *Olympics* is a dummy variable with 0 indicating there's no

constant event for this martial art style in the Olympics and 1 for the opposite. *Competition* equals to the sum of the product of how many years a competition has been active and how many nations have achieved medals in that competition for all renowned international competitions for that specific martial art style, the same for *Performance*.

Given that the random error term contains all variables that are correlated with the dependent variables that are not listed as independent variables, all the control variables are linked to both the independent variables and the dependent variables. *Olympics* would affect the *Popularity* directly and also affect *Movien*, *Themetime* given that more modern Olympics topics can be featured, and also affecting *Mainstyle* as it would be more likely to be featured as the mainstyle when it's own events is featured. *Competition* and *Performance* will both affect *Popularity* directly and the *Combatvalue* given that the style with more sparring events tend to be featured with its combat value and the one with more performance events can be featured otherwise. They can also affect *Themetime* for the same reason as *Olympics*. *Game* can affect *popularity*, *Boxoffice* and even *Combatvalue* as action games usually exaggerate the movements and tend to affect action designs in films.

All the control variables can be seen in the appendix.

5 Model Construction and justification

A Multivariate Linear Regression model is constructed to properly evaluate the relationship between the independent variables and dependent variables through the coefficients of relevant terms. The specific model is constructed as below

$$\begin{aligned} Popularity =& \alpha + \sum \beta_m \times [Movien + Movien \times (Combatvalue + Mainstyle) + Boxoffice \\ &+ Nationality + Themetime + TimeIndex] + \sum \lambda \times ControlVariables + \epsilon \end{aligned} \tag{1}$$

In equation 1, α is the constant term whereas ϵ is the random error term containing all the variables that are correlated with *Popularity* but not listed out. *Movien* are timed with *Combatvalue* and *Mainstyle* as they have interaction effects with *Movien*. For example, the average Combatvalue for all the movies for a specific martial art style would be insignificant for the popularity as it is hard to be detected and would be more significant and easier to be detected when it gets added up reflected in *Movien*, same reasoning for the other two explanatory variables.

6 Results and Discussions

6.1 Regression Result

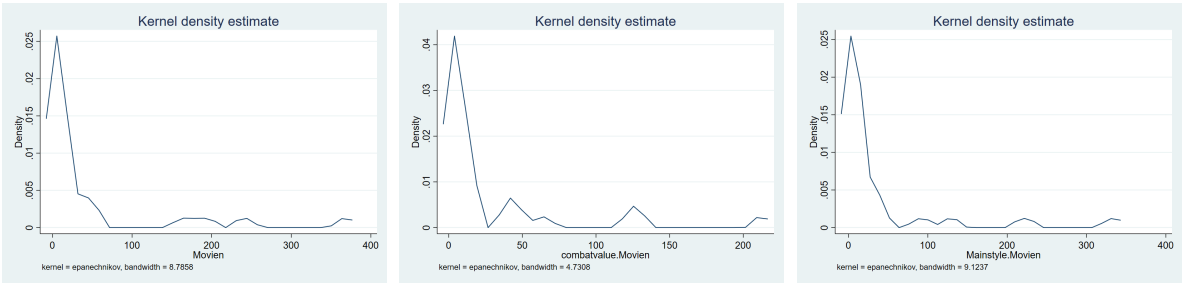
The coefficients of Robust multiple linear regression for the model constructed are listed below in figure 2. Robust regression is run to avoid heteroscedasticity in the regression model which would, in turn, lead to invalidity of the regression model and the result. It can be clearly seen that the variable term with a coefficient significant at the significance level of 0.05 is $Competition$, terms with a coefficient significant at the significance level of 0.01 are $Combatvalue \times Movien$, $Game$,$Performance$, and *Movien*. The rest of the variables all have a coefficient that is not considered statistically significant at any significance level.

VARIABLES	(1) Coefficient
Movien	-28.561***
	(-4.41)
Nationality	-163.857
	(-0.56)
Boxoffice	-0.046
	(-0.02)
combatvalueMovien	47.560***
	(10.96)
Themetime	-53.316
	(-1.24)
MainstyleMovien	5.374
	(1.08)
TimeIndex	-97.434
	(-1.21)
olympics	-888.060
	(-1.64)
competition	0.176**
	(2.71)
performance	3.037***
	(17.88)
game	6.682***
	(3.92)
Constant	417.027
	(1.61)
Observations	30
R-squared	0.929

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 2: Regression result

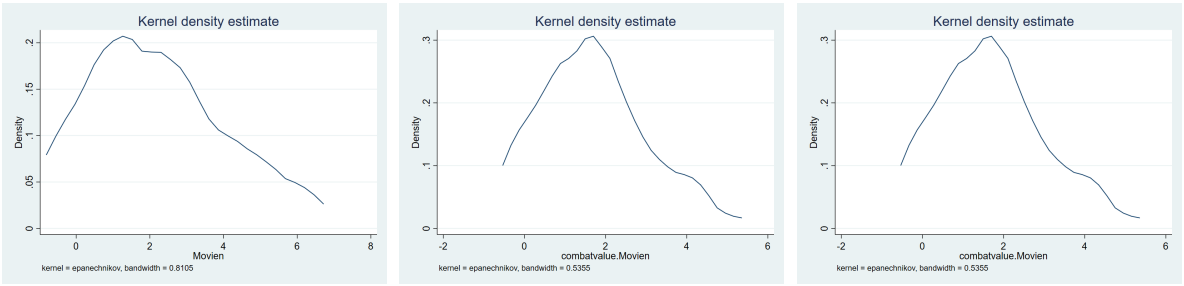
A notable problem reflects in the coefficient of *Movien*, movie number has no reasonable explanations to be having negative correlation with the popularity. The case that lots of martial art styles with little popularity are featured in very few films is also noticed during the collection of data. This phenomenon is identified to be caused by the unbalanced distribution of *Movien*.



(a) (b) (c)

Figure 3: Kernel density estimate for *Movien, Combatvalue*×*Movien,Mainstyel*×*Movien*

From figure 3, the kernel density estimation shows the high Kurtosis of *Movien* and all the terms involving *Movien*. A large proportion of the data are distributed at rather lower values and causing the mispredictions of the coefficients and its sign. To address this problem, *Movien* is changed to the ln(number of movies). Consequently, the data will be more balanced and well-distributed as shown in figure 4 below.



(a) (b) (c)

Figure 4: Kernel density estimate for *Movien, Combatvalue*×*Movien,Mainstyel*×*Movien* after processing

As figure 4 gives promising result, the robust regression is done again with new *Movien* is done and the result is shown below.

VARIABLES	(1) Coefficint
Movien	-934.755
	(-1.37)
Nationality	404.558
	(0.87)
Boxoffice	-3.511
	(-1.38)
combatvalueMovien	1,777.600**
	(2.40)
Themetime	-74.000
	(-1.01)
MainstyleMovien	-109.808
	(-0.32)
TimeIndex	-132.610
	(-1.06)
olympics	-834.459
	(-1.64)
competition	0.201**
	(2.56)
performance	2.386***
	(4.15)
game	13.774
	(1.70)
Constant	-467.171
	(-0.89)
Observations	30
R-squared	0.860

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 5: Regression result

The regression shows a new result with the coefficient for *Movien* insignificant which means 0. It can be clearly seen that the variable term with a coefficient significant at the significance level of 0.05 are *Competition* and $Combatvalue \times Movien$, terms with a coefficient significant at the significance level of 0.01 is $performance$.

6.2 Evaluation

The principle for evaluation is that a model can be considered reliable if 1)there's no heteroscedasticity existing in the model. 2)there's no serious endogeneity on the explanatory variables. 3)there's a

relatively small VIF that ensures the lack of multicollinearity.

Data collected in this study is typical cross-sectional data that is vulnerable to suffering from heteroscedasticity problems. As the Robust Regression is already the regression model that has been processed with Huber-white standard errors, the model should theoretically possess no heteroscedasticity problems. However, a test is performed to confirm the absence of heteroscedasticity problems. The white test would be the optimal choice as the White test can be used to examine the heteroscedasticity in all kinds of forms.

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity
chi2(29) = 30.00
Prob > chi2 = 0.4140

Cameron & Trivedi's decomposition of IM-test

chi2	df	p
30.000	29	0.414
7.370	11	0.768
0.780	1	0.378
38.150	41	0.598

Figure 6: White test result

The result shows a Prob=0.414>0.05 which means the acceptance of null hypothesis H0 of homoskedasticity and disproves the claim of the model attaining heteroscedasticity.

As the heteroscedasticity problems are solved, in order to ensure the validity and consistency of the regression model and coefficient, we also have to consider the possibility of independent variables having endogeneity in respect to the random error terms. Because of fact that ensuring the exogeneity of all the variables, including the control variables would be difficult and also unnecessary. The endogeneity of only the explanatory variables will be the focus of our evaluation.

Therefore, 2SLS regression with explanatory variables as instrumented variables is conducted to be compared with the original regression result. To conduct 2SLS regression, instrumental variables are selected fulfilling the following principle: An instrumental variable can only affect the *Popularity* through affecting an explanatory variable, given the following requirements of an instrumental variable (z_i)

$$Cov(z_i, \epsilon) = 0$$
$$Cov(z_i, x_i) \neq 0$$

The following instrumental variables are collected and used in the 2SLS regression.

Variable	Variable Description	Range
History	the number of centuries betweem the 21st century and the time an existing martial art style has formed in	N or 0
CountryGPI	Global Power Index of different origin countries for a specifc style of martial art	[0,+∞)
CountrySP	Soft Power Index of different origin countries for a specific style of martial art	[0,+∞)
Branches	evaluate the number of substyles of a specific martial art style	N
Combativity	evaluate the frequency a specific martial art style appear in international inter-style events	[0,1]
Rating	the average rating of the movies featuring a martial art style	[0,5]
HED	how many decades from 2020s has the origin country reached its highest GDP world rank	[0,+∞)

Variable	Variable Description	Range
OrganIndex	evaluate the scale of the largest international organization of a specific style of martial art	[0,+)

Table 4:Instrumental Variables

In the instrumental variables in table 4, *CountryGPI* is taken from Global Power Index 2020 (PANGAEA WIRE GROUP 2020). All the rating information is from LetterBox, as for HED data are from data from World Bank.

OrganIndex takes the form of the product of the number of member countries and how long has the organization been founded.

CountrySP is taken from Global Soft Power Index 2020(BrandFinance 2020).*Combativity* for martial art style j is taken with the following equation:

$$Combativity_j = \sum_{i=1}^{Num} \frac{Viewership_i}{\sum_{i=1}^{Num} Viewership_i} \times \frac{Occurance_j}{TotalOccurance}$$

in which Num stands for the total number of large cross-style combat competitions that was taken in to consideration, $Occurance_j$ stands for how many people have used martial art style j in one competition and TotalOccurance is the sum of that value for all collected martial art styles. This ensures the *Combativity* for all the martial art styles add up to 1 and serve the function of showing the rate at which one martial art style is represented in international events. The component data for *Combativity* can be seen in the appendix.

The planned relationship between instrumented variables and the instrumental variables are as following:

$$\begin{aligned} Themetime &= \gamma_1 \times History + \epsilon_1 \\ Nationality &= \gamma_2 \times Branches + \epsilon_2 \\ Movien &= \gamma_3 \times CountryGPI + \gamma_4 \times CountrySP + \epsilon_3 \\ Combatvalue &= \gamma_5 \times Combativity + \epsilon_4 \\ TimeIndex &= \gamma_6 \times HED + \epsilon_5 \\ Mainstyle &= \gamma_7 \times OrganIndex + \epsilon_6 \\ Boxoffice &= \gamma_8 \times Rating + \epsilon_7 \end{aligned}$$

All the instrumental variables' data can be seen in the appendix.

The 2SLS IV robust regression is then run according to the information above with the result below

Instrumental variables (2SLS) regression

popularity	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Movien	-1812.543	1235.229	-1.47	.142	-4233.548	608.462	
Nationality	-481.576	1735.356	-0.28	.781	-3882.813	2919.66	
Boxoffice	-11.63	6.988	-1.66	.096	-25.326	2.066	*
combatvalueMovien	2733.217	811.203	3.37	.001	1143.289	4323.145	***
Themetime	-96.619	157.536	-0.61	.54	-405.383	212.145	
MainstyleMovien	-4.113	912.624	-0.00	.996	-1792.824	1784.598	
TimeIndex	62.356	310.28	0.20	.841	-545.783	670.495	
olympics	-1018.311	1271.763	-0.80	.423	-3510.921	1474.299	
competition	.144	.094	1.52	.128	-.041	.329	
performance	2.583	.583	4.43	0	1.44	3.725	***
game	24.39	8.624	2.83	.005	7.488	41.292	***
Constant	115.296	1598.628	0.07	.943	-3017.958	3248.549	

Mean dependent var	882.825	SD dependent var	1425.367
R-squared	0.754	Number of obs	30.000
Chi-square	292.456	Prob > chi2	0.000

*** p<.01, ** p<.05, * p<.1

Figure 7: 2SLS regression result

As the robust regressions are used, the Durbin Wu-Hausman test is ran using the result of Robust regression and 2SLS regression. The comparison between those two regression models can be used to determine whether there is an endogeneity problem in the original robust regression model. The Durbin Wu-Hausman test result is listed below.

Tests of endogeneity
Ho: variables are exogenous
Robust score chi2(7) = 8.29224 (p = 0.3075)
Robust regression F(7,11) = 1.34233 (p = 0.3178)

Figure 8: Durbin Wu-Hausman test result

From figure 8, it can be told that the p-value is larger than 0.05 and thus strongly confirming the original hypothesis of the explanatory variables are exogenous, disproving the existence of a serious endogeneity problem thus proving the reliability of the regression model. Whether the instrumental variables are correlated with the random error term is then tested with the overidentification test and provided the following result

Test of overidentifying restrictions:
Score chi2(1) = .374349 (p = 0.5406)

Figure 9: Overidentification test result

The result above shows a p-value of 0.5406 which strongly confirms the H0 of all the instrumental variables are exogenetic which means no correlations with the random error term and thus confirm the validity of the instrumental variables used.

As the exogeneity of the regression model is proven, the possibility of multicollinearity within the regression model needs to be considered. To examine the multicollinearity of the regression model, we calculate the variance inflation factor(VIF) for all variables in the regression model. VIF can be depicted through the following equation

$$VIF_m = \frac{1}{1 - R^2_{1-km}}$$

in which R^2_{1-km} is the R-squared value for the regression model taking the m^{th} variable as the dependent variable and the rest k-1 ones as independent variables. Therefore, variables with a larger VIF will be more related to other variables and cause multicollinearity.

	VIF	1/VIF
Movien	83.470	0.012
CombatValue×Movien	52.950	0.019
Mainstyle×Movien	35.720	0.028
game	20.580	0.049
Olympics	3.110	0.321
Competition	2.750	0.364
Nationality	2.520	0.397
Boxoffice	2.460	0.407
Performance	2.460	0.407
TimeIndex	2.020	0.495
Themetime	1.620	0.618
Average VIF	19.060	

Figure 10: Variance inflation factors for all variables

Figure 10 shows that the term of *Mainstyle×Movien, Combatvalue×Movien, Movien, and game* all possess VIFs higher than 10 that indicate significant multicollinearity for these variables which would increase the variance and leads to mis-prediction of the coefficients for the variables mentioned above. To eliminate the effect of multicollinearity upon the regression result, stepwise regression is run with a pe value of 0.3 to specify the p value to be lower than 0.3 to be considered in the regression. This will eliminate redundant variables that are not significant and in return decrease the variance inflation factor for the rest of the variables which will provide an accurate coefficient for the variables above.

The stepwise regression ultimately leaves six variables which are *Movien×Mainstyle*, *Movien×Combatvalue*, *Competiton*, *Nationality*, *Themetime* and *Performance*.The regression result is shown in figure 11.

The result shows that *Competiton* and *Performance* are significant at 0.01 significance level, *Themetime* and *Movien×Combatvalue* are significant at 0.05 level.

VARIABLES	(1) Coef.
MainstyleMovien	38.785
	(0.18)
competition	0.168***
	(3.29)
performance	1.975***
	(3.15)
Themetime	-148.980**
	(-2.13)
combatvalueMovien	525.858**
	(2.15)
Nationality	747.159
	(1.61)
Constant	-904.858*
	(-1.98)
Observations	30
R-squared	0.813

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 11: Regression of variables with p value ≤ 0.01

After stepwise regression, the VIF are calculated again and shows the following result

	VIF	1/VIF
Mainstyle×Movien	6.630	0.151
Combatstyle×Movien	6.040	0.166
Performance	2.110	0.474
Competition	1.430	0.699
Nationality	1.290	0.776
Themetime	1.150	0.873
Average VIF	3.110	

Figure 12: VIF for with p value ≤ 0.01

From figure 12, it can be seen that no VIF is larger than 10 after excluding some variables that are not significant which solves the existence of significant multicollinearity for the coefficients in the regression models and indicates accuracy in the coefficient outcome. However, lots of control variables are excluded. Endogeneity might reoccur and needs to be tested again through the Durbin Wu-Hausman test.

Tests of endogeneity
Ho: variables are exogenous
Durbin (score) chi2(5) = 2.0752 (p = 0.8386)
Wu-Hausman F(5,17) = .252668 (p = 0.9326)

Figure 13: Durbin Wu-Hausman test result after stepwise regression

From the result above, it can be seen that even with the exclusions of variables after stepwise regression, the null hypothesis——exogeneity of the explanatory variables is strongly confirmed and the instrumental variables for the explanatory variables are also exogenetic according to the overidentification test conducted before.

Therefore, all the criteria are met and the model can be considered reliable. The final model with terms with significant coefficients is the following:

$$\begin{aligned} Popularity = 525.858 \times Movien \times Combatvalue - 148.980 \times Themetime \\ + 0.168 \times Competition + 1.975 \times Performance - 904.858 \end{aligned} \tag{2}$$

The *Popularity-Themetime-Movien* $\times$ *Combatvalue* graph is shown below with the samples (Processed Popularity is the $popularity - (0.168 \times Competition + 1.975 \times Performance - 904.858)$)

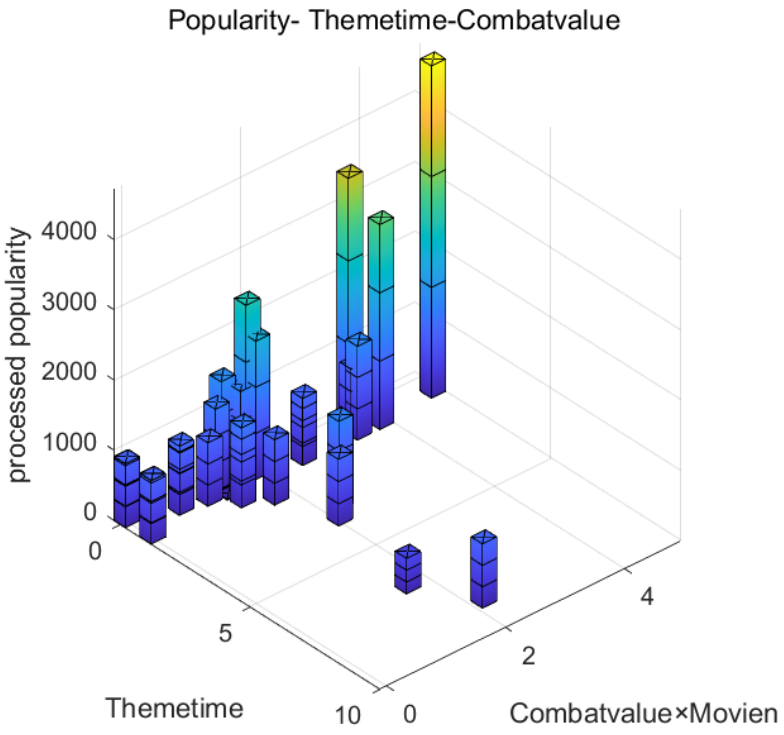


Figure 14: Processed Popularity-Themetime-*Movien* $\times$ *Combatvalue*

It can be seen that there are obvious descending trends among both axis which proved the validity of the model in a sense.

6.3 Limitaions

Limitations in this study are majorly reflected in the following aspects:

1. Data collection is done manually and may not be sophisticated enough. This will have to be resolved by the construction of a scientific martial art database which is currently lacking nowadays. The lack of martial art database surely has put large burdens in front of scientific researches concerning martial art.

2.The sample size is not big enough. This is due to the status quo at which there's an obvious phenomenon of an unbalanced distribution of popularity for different martial art styles.

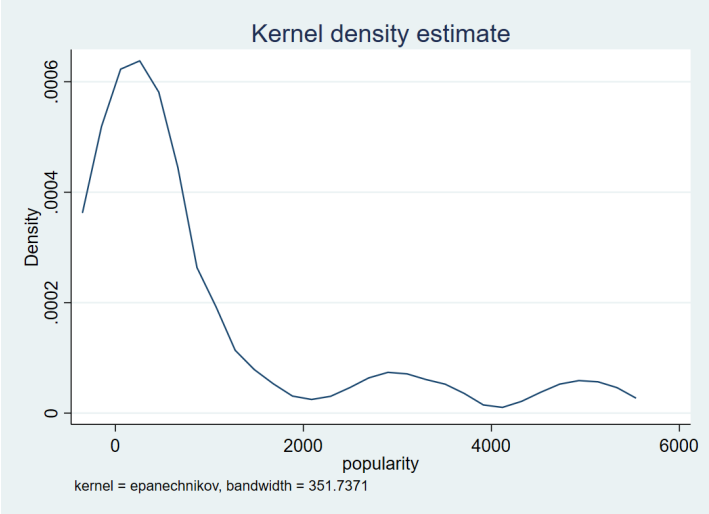


Figure 15: kernel density estimate for *Popularity*

With figure 15's information, it can be seen that the Kurtosis of popularity cannot afford more unbalanced distribution which will lead to relatively large kurtosis and eventually affect the reliability of the value of the coefficients and even the plus or minus sign of the coefficients. This is a major limitation of the sample itself.

7 Conclusions

From the robust regression and stepwise regression results presented above in equation 2, some conclusions can be drawn.

Firstly, *Combatvalue* of one martial art style presented in movies has a significant impact on the popularity of that martial art style with the interaction of the number of movies. The more movies with high *Combatvalue*, the higher the possibility. This proves the core focus with respect to a martial art style in action films eventually falls on combats. Even though the audience is engaged in an activity of watching artistically modified martial arts, it's still optimal for them to see martial arts modified in, at least seemingly, a realistic way. Martial art styles with lots of movies with high *Combatvalue* will be covered with an impression of being well developed in combats which is the essential and the root for martial arts and thus, attain great popularity. This conclusion proves the value of the work done by action directors that tries to depict martial art styles in realistic combats clearly like Donnie Yen, Woo Ping Yuen, and so on. On the other hand, action designs that are largely separated from realistic combats themselves will not help the popularity of a martial art though it still possesses artistic values.

Secondly, *Themetime* seem to be negatively related to the popularity of a martial art style with interactions with movie numbers. This indicates that people are more likely to be influenced by action

movies taken with theme times near the current century given that the plots are developing in a more familiar way. Additionally, martial arts expressed in the action movies with small *Themetime* tend to be more realistic while those featuring ancient martial arts tend to be using wires frequently and thus lack *CombatValue* (e.x. Flying Swords of Dragon Gate 2011). Whether there is a correlation between *Themetime* and *CombatValue* requires further investigation

Thirdly, *Boxoffice*, *Nationality*, *Mainstyle*, *Movinen* and *TimeIndex* don't seem to have strong correlation with the popularity of a style of martial art. The real quality and contents of the action design of martial art styles themselves seem to substantially outweigh the descriptive information of movies in case of the impact on the popularity of a martial art style.

All in all, the quality of action designs in the featured actions is the key for promoting corresponding martial art styles which are largely lacking in lots of the modern 'action movies', especially kung fu films. Actors with proper training in fighting scenes would be substantially important in such movies and are nearly the prerequisite of high-quality action designs which has decreased by a large extent since the decline of Hong Kong movie industry(HKU 2016). It can be detected how various Wuxia movies with no combat value whatsoever and only flying around is now more and more common and slowly turning into the main trend in Chinese kung fu movies. While these movies are reasonable on their own, it is still necessary to bring action movies with high-quality action designs back to the major's view to help to promote different martial art styles to a larger population and preserve this cultural heritage.

Appendix

Explanatory and explained variables

	Movien	Nationality	Boxoffice	combatvalue	Themetime	Mainstyle	TimeIndex	popularity
mma	51	1.00	12.9	0.922	0.294	0.765	1.961	3339.584
Kung fu	368	0.90	21.3	0.345	3.538	0.908	4.052	2458.244
judo	4	0.50	43.1	1.000	0.750	1.000	3.500	1278.278
karate	165	0.22	29.0	1.000	0.945	0.500	3.461	3074.713
kick-boxing	46	0.85	17.3	0.880	0.696	0.921	2.913	450.694
wrestling	28	1.00	19.4	0.758	0.286	0.558	1.929	4773.483
Muay Thai	43	0.63	54.2	0.913	0.930	0.739	1.767	693.900
Taekwondo	16	0.50	0.9	0.571	0.875	1.000	3.000	1064.768
Hapkido	7	0.43	1.2	0.930	1.286	0.767	3.714	71.970
Capoeira	7	0.43	5.4	1.000	0.286	0.667	2.143	543.305
Kalaripayattu	21	0.95	10.4	0.938	8.762	0.813	2.857	9.892
sumo	9	0.78	0.0	0.714	0.556	0.286	2.222	1136.201
Ninjitsu	193	0.47	28.1	0.714	1.751	0.429	3.212	23.861
krav-maga	5	0.00	146.6	0.714	0.000	0.810	0.800	496.954
BJJ	6	0.50	4.4	0.889	0.333	1.000	1.500	512.778
boxing	241	1.00	12.4	0.332	0.747	0.668	4.456	5188.885
savate	1	1.00	0.0	1.000	1.000	0.000	3.000	37.736
silat	7	1.00	2.6	1.000	1.286	1.000	2.286	302.151
Jeet Kune do	16	0.81	71.3	0.714	0.875	1.000	3.563	60.423
Lethwei	2	0.50	0.0	0.938	0.500	0.750	1.000	4.503
Aikido	12	0.00	42.7	1.000	3.000	1.000	2.667	471.573
Bartitsu	1	1.00	0.0	3.667	1.000	2.667	1.000	2.458
sambo	1	1.00	0.0	1.000	0.000	1.000	0.000	188.387
Muay Lao	1	1.00	0.0	1.000	1.000	0.000	1.000	0.550
Arnis	3	1.00	14.3	1.000	0.667	0.667	2.667	54.489
kenpo	7	0.86	14.9	6.000	7.000	5.000	3.000	122.887
vale tudo	2	1.00	0.0	0.895	0.500	0.789	1.000	113.682
bokh	2	1.00	0.0	1.000	0.500	1.000	1.000	2.585
Angampora	1	1.00	0.0	1.000	0.000	1.000	1.000	0.870
dambe	1	1.00	0.3	1.000	0.000	0.500	3.000	4.955

Movies at different time periods

	movien	30s	40s	50s	60s	70s	80s	90s	00s	10s	20s
mma	51	0	0	0	0	0	2	8	27	14	0
Kung fu	368	0	0	0	19	164	98	27	27	30	3
sambo	1	0	0	0	0	0	0	0	0	0	1
judo	4	0	1	0	0	0	0	1	1	1	0
boxing	241	22	36	22	5	20	27	20	48	40	1
karate	165	0	1	0	2	39	42	47	15	17	2
kick-boxing	46	0	0	0	0	1	7	30	3	5	0
wrestling	28	0	0	0	0	1	4	1	8	14	0
Muay Thai	43	0	0	0	0	1	4	1	15	22	0
brazilian Jiu-jitsu	6	0	0	0	0	0	0	1	2	2	1

	movien	30s	40s	50s	60s	70s	80s	90s	00s	10s	20s
Taekwondo	16	0	0	0	0	2	2	7	4	1	0
Hapkido	7	0	0	0	0	2	1	4	0	0	0
Capoeira	7	0	0	0	0	0	0	2	4	1	0
Kalaripayattu	21	0	0	0	3	3	0	4	4	7	0
sumo	9	0	0	1	0	0	0	1	3	4	0
Ninjutsu	193	0	0	0	18	15	70	23	25	38	4
krav-maga	5	0	0	0	0	0	0	0	0	4	1
savate	1	0	0	0	0	0	0	1	0	0	0
silat	7	1	0	0	0	0	0	0	1	5	0
Jeet Kune do	16	0	0	0	0	6	3	4	0	3	0
Lethwei	2	0	0	0	0	0	0	0	0	2	0
Aikido	12	0	0	0	0	0	1	8	1	2	0
Bartitsu	1	0	0	0	0	0	0	0	0	1	0
Muay Lao	1	0	0	0	0	0	0	0	0	1	0
Arnis	3	0	0	0	0	0	1	1	0	1	0
kenpo	7	0	0	0	0	0	0	7	0	0	0
fencing	19	1	1	2	1	1	1	5	5	2	0
vale tudo	2	0	0	0	0	0	0	0	0	2	0
Angampora	1	0	0	0	0	0	0	0	0	1	0
bokh	2	0	0	0	0	0	0	0	0	2	0

Control Variables

	olympics	competition	performance	game
mma	0	4281	0	22
Kung fu	0	1108	1128	268
judo	1	13071	0	8
karate	0	6320	0	44
kick-boxing	0	2380	0	4
wrestling	1	5139	1205	138
Muay Thai	1	1566	0	15
Taekwondo	1	4216	228	5
Hapkido	0	0	0	2
Capoeira	0	130	130	3
Kalaripayattu	0	0	0	0
sumo	0	441	0	2
Ninjitsu	0	0	0	287
krav-maga	0	0	0	0
brazilian Jiu-jitsu	0	907	0	3
boxing	1	8108	0	103
savate	0	450	0	4
silat	0	414	0	3
Jeet Kune do	0	52	0	6
Lethwei	0	126	0	2
Aikido	0	676	0	10
Bartitsu	0	0	0	1
sambo	0	1276	0	8
Muay Lao	0	11	0	0
Arnis	0	342	0	3

	olympics	competition	performance	game
kenpo	0	454	225	3
vale tudo	0	210	0	2
bokh	0	15	0	0
Angampora	0	0	0	1
dambe	0	0	0	1

Instrumental Variables

	history	countryGPI	countrysp	branches	combativity	rating	HED	OrganIndex
mma	2	0.53	55.6	1	0.223	2.8	0.667	1080
Kung fu	40	1.00	58.7	129	0.049	3.0	0.000	4805
judo	2	0.15	60.2	6	0.008	2.9	1.000	14000
karate	4	0.15	60.2	11	0.025	2.9	1.000	6138
kick-boxing	1	0.36	55.0	13	0.069	2.9	0.333	3538
wrestling	1	0.57	59.1	21	0.049	2.7	1.667	18748
Muay Thai	5	0.00	37.6	5	0.047	3.0	0.000	3640
Taekwondo	20	0.00	48.3	11	0.148	2.9	1.000	9984
Hapkido	1	0.00	48.3	4	0.063	3.4	1.000	2000
Capoeira	5	0.51	39.4	3	0.191	2.6	1.000	495
Kalaripayattu	30	0.49	41.6	3	0.041	2.2	0.000	0
sumo	13	0.15	60.2	6	0.022	3.1	1.000	2523
Ninjitsu	8	0.15	60.2	19	0.001	2.9	1.000	870
krav-maga	1	0.11	42.6	3	0.013	3.5	1.000	2064
brazilian Jiu-jitsu	1	0.51	39.4	25	0.000	3.5	1.000	2109
boxing	27	0.47	55.5	1	0.000	2.6	0.667	15500
savate	2	0.18	58.5	1	0.000	2.9	1.000	2268
silat	8	0.00	33.4	150	0.001	3.3	0.000	2706
Jeet Kune do	1	1.00	58.7	2	0.003	3.3	0.000	1537
Lethwei	16	0.00	20.0	1	0.007	0.0	0.000	128
Aikido	7	0.15	60.2	16	0.009	2.3	1.000	2632
Bartitsu	1	0.19	61.8	1	0.009	1.4	0.000	0
sambo	1	0.65	51.0	2	0.001	1.3	4.000	3525
Muay Lao	5	0.00	19.0	1	0.000	0.0	0.000	0
Arnis	6	0.00	32.5	3	0.000	3.4	2.000	1131
kenpo	1	0.92	67.1	5	0.001	3.0	0.000	1900
vale tudo	1	0.51	39.4	1	0.015	2.2	1.000	210
bokh	23	0.00	18.3	5	0.000	0.0	0.000	0
Angampora	25	0.00	19.2	64	0.000	0.0	0.000	0
dambe	20	0.10	28.8	1	0.001	3.5	0.000	0

Popularity data

Popularity data are shown in https://github.com/w0212/Popularity due to length limitaions.

Combativity data

	UFC	Glory of heroes	K-1	glory	One Championship	Bellator
MMA	160	29	78	22	77	35
Kung fu	3	21	21	10	47	7
sambo	3	0	4	1	12	0
judo	6	0	22	1	44	7
boxing	23	8	56	12	76	0
karate	8	15	63	7	55	7
brazilian jiu-jitsu	13	1	21	2	69	49
kick-boxing	17	22	142	83	105	14
wrestling	35	2	35	0	63	42
Muay Thai	31	29	96	71	190	14
Jiu-jitsu	32	10	3	2	0	14
Taekwondo	2	5	24	2	35	0
Hapkido	1	0	2	0	0	0
Capoeira	5	3	7	4	7	0
Kalaripayattu	0	0	0	0	0	0
sumo	0	0	0	0	0	0
Ninjutsu	0	0	0	0	0	0
krav-maga	1	0	0	0	0	0
savate	0	0	23	3	2	0
silat	3	0	2	2	7	0
Jeet Kune do	5	3	0	1	3	0
Lethwei	0	2	1	0	17	0
Aikido	0	0	0	0	2	0
Bartitsu	0	0	0	0	0	0
Muay Lao	0	0	1	0	0	0
Arnis	1	0	0	0	1	0
kenpo	4	0	21	2	1	0
vale tudo	10	5	0	2	0	0
Angampora	0	0	0	0	0	0
bokh	0	0	0	0	0	0
dambe	1	0	0	0	0	0
Total	364	155	622	227	813	189
viewership(million)	6.1	2.5	0.0	3.5	6.1	0.3
Influence Factor	0.33	0.14	0.00	0.19	0.33	0.02

Reference

(1)“Ratings, Reviews, and Where to Watch the Best Movies & TV Shows.” IMDb, IMDb.com, www.imdb.com/.

(2)Letterboxd • Social Film Discovery. letterboxd.com/.

(3)“Credits, Top Critics, Reviews, Videos and Screenshots.” IGDB.com, www.igdb.com/.

(4)Google Trends, Google, trends.google.com/trends/?geo=US.

(5)Farquhar, Mary. “Jet Li: ‘Wushu Master’ in Sport and Film.” Celebrity in China, 2010, pp. 103–123., doi:10.5790/hongkong/9789622090873.003.0006.

(6)Wong, Suet-lan. “Hong Kong Cinema Made International : the Action Cinema of Bruce Lee and Jackie Chan.” doi:10.5353/th_b3195176.

(7)Bolelli, Daniele. “How Gladiatorial Movies and Martial Arts Cinema Influenced the Development of The Ultimate Fighting Championship.” JOMEC Journal, no. 5, 2014, doi:10.18573/j.2014.10265.

(8)Dana White, Personal interview 5 March 2008

(9)Li Yuan, Wang Gang, Zhu Ruiqi. Formation and reflection of negative image of Chinese Martial Arts [J]. Journal of Beijing Sport University, 2013, 36 (9): 37 Advances in Social Science, Education and Humanities Research, volume 83

(10)Zhang Zhi Bai, interview. By Junfei Sun. 15 Jan 2021

(11)Cheng Yu, interview. By Junfei Sun. 17 Jan 2021

(12)Ping Zhang, interview. By junfei Sun. 21 Jan 2021

(13)Jian Shi, interview. By Junfei Sun. 21 Jan 2021

(14)“Homepage.” UFC.COM, www.ufc.com/.

(15)“ONE Championship - The Home Of Martial Arts.” ONE Championship – The Home Of Martial Arts, www.onefc.com/.

(16)K. Lee, Alexander, et al. “Bellator.” Bellator - MMA Fighting, www.mmafighting.com/bellator.

(17)“1 WORLD CHAMPIONS - K-1 WORLD.” K, 22 Mar. 2017, k-1world.com/all-k-1-champions/.

(18)Glory Kickboxing, glorykickboxing.com/.

(19)“About IWUF.” IWUF, www.iwuf.org/about-iwuf/.

(20)IJF.org - International Judo Federation, www.ijf.org/.

(21)“World Karate Federation.” WKF, www.wkf.net/.

(22)IKF Kickboxing, www.ikfkickboxing.com/.

(23)International Federation of Muaythai Associations. muaythai.sport/.

(24)United World Wrestling, uww.org/.

(25)INTERNATIONAL HAPKIDO FEDERATION, www.hapkido.or.kr/.

(26)World Capoeira Federation, www.capoeira.ws/.

(27)"World Taekwondo." Worldtaekwondo.org, www.worldtaekwondo.org/index.html.

(28)International Ninja Federation / , 10 Mar. 2017, www.international-ninja-federation.com/.

(29)"IKMF Krav Maga." IKMF, 8 June 2021, kravmaga-ikmf.com/.

(30)"International Sumo Federation." International Sumo Federation, ifs-sumo.org/.

(31)"WBA Official Site." WORLD BOXING ASSOCIATION, www.wbaboxing.com/.

(32)"International Brazilian Jiu-Jitsu Federation." IBJJF, ibjjf.com/.

(33)"Fédération Française De Savate Boxe Française Et D.A. - FISAV." Fédération Française De Savate Boxe Française Et D.A. - Accueil, www.ffsavate.com/fisav.html.

(34)Home - INTERNATIONAL JEET KUNE DO KIM FEDERATION. jkdkim.com/.

(35)"Linear." International Pencak Silat Federation (PERSILAT), www.epsf.net/persilat.html.

(36)"International Aikido Federation (IAF)." International Testing Agency, 4 May 2021, ita.sport /partner/international-aikido-federation-iaf/.

(37)International SAMBO Federation (FIAS), sambo.sport/.

(38)"International Lethwei Federation Japan - ILFJ: Lethwei Promoter." Tapology, www.tapology. com/fightcenter/promotions/3377-international-lethwei-federation-japan-ilfj.

(39)"International Arnis Federation." Americanarnisassociationorg, americanarnisassociation.org /history/international-arnis-federation/.

(40)"International Vale Tudo Championships." EDGE MMA, mixedmartialarts.fandom.com /wiki/International_Vale_Tudo_Championships.

(41)Home." INTERNATIONAL KEMPO FEDERATION, olympickempo.com/.

(42)The Cinema Development of Hong Kong: Is Hong Kong Still a ... sociology.hku.hk/wp-content/uploads/gravity_forms/15-662cc63985444226b987fafadce3ea50/preliminary-research_HK-Cinema-Development.pdf.