Summer 2016

Extension of the Gravity Model: A Risk Integrated Approach Towards the Impact Analysis of Mega Sports Events on Inbound Tourist Arrivals

Abdul Sami Stanekzai
Old Dominion University, astan002@odu.edu

EXTENSION OF THE GRAVITY MODEL: A RISK INTEGRATED APPROACH TOWARDS THE IMPACT ANALYSIS OF MEGA SPORTS EVENTS ON INBOUND TOURIST ARRIVALS

by

Abdul Sami Stanekzai
B.Sc. December 2011, Kabul University, Afghanistan

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

ENGINEERING MANAGEMENT

OLD DOMINION UNIVERSITY
August 2016

Approved by:

Ghaith Rabadi (Director)

T. Steven Cotter (Member)

Mamadou Seck (Member)

ABSTRACT

EXTENSION OF THE GRAVITY MODEL: A RISK INTEGRATED APPROACH TOWARDS THE IMPACT ANALYSIS OF MEGA SPORTS EVENTS ON INBOUND TOURIST ARRIVALS

Abdul Sami Stanekzai
Old Dominion University
Director: Dr. Ghaith Rabadi

Mega sports events such as the Olympics and the FIFA World Cup are highly attended and countries compete ferociously to host such events due to their perceived long term positive effects. Inbound tourist forecasting is an important aspect of the hosting decision both for the organizing committees and the hosting nations. As a precaution against letting the event fall into incapable hands, which can lead into chaos and mismanagement, measures needs to be taken. Capacity evaluation of interested parties is possible, only by knowing the number of people who may attend the event. However, due to the infrequent occurrence of such events, it is not straightforward to do so.

In this thesis, focus is made on a regression model known as the Gravity Model to predict the number of inbound tourists between pairs of countries. A large set of quantitative and quantified-qualitative factors having impact on the touristic behavior of people is studied and models from the literature are validated using this data. The results are discussed and important suggestions are made. Moreover, the Gravity Model is extended and new predictors are introduced accounting for significant aspects of tourism in general and mega events in particular. The new model will give more accurate results potentially.

This thesis is dedicated to two of the most important people in my life; my father Abdul Jalil Stanekzai, and my loving wife Tahmina Tabee Stanekzai. The credit of all my education from elementary school, to my undergraduate studies and now master's degree goes to my father. Without his endless encouragement and nonstop support I wouldn't have been able to accomplish any of this. The love and support of my wife along with her patience and understanding have also been vital in successful completion of my research and the degree program as a whole.

ACKNOWLEDGEMENTS

There are a number of people who have contributed to successful completion of this thesis. First of all, I would like to extend my sincere gratitude to my advisor, Dr. Ghaith Rabadi, for his invaluable guidance and extensive support throughout the process. Dr. Rabadi always remained positive when I dealt with the hardships on this journey. His wisdom, patience, and encouragement has been substantial in successful completion of this research.

I would also like extend my sincere appreciations to the committee members. Dr. Steven Cotter's expertise and willingness to help even on his weekends require special recognition. This thesis wouldn't be complete without recognizing the insightful questions of Dr. Mamadou Seck which broadened my perspective and opened new areas to explore. I would also like to thank the U.S. Department of State and the Fulbright scholarship program for funding my degree program in the first place.

Last but not least, thanks to my friends, Masoor Ahmed for always being there for me when I needed him, and Ying Thaviphoke for allowing me to use his computer on several occasions.

NOMENCLATURE

α	Alpha, Statistical Significance Level
β	Beta, Coefficients Estimate
CEPII	Centre d'Etudes Prospectives et d'Informations Internationales
CIA	Central Intelligence Agency
C.I.F	Cost, Insurance and Freight
CU	Common Currency
CWC	Cricket World Cup
δ	Delta, Origin Fixed Effects
DOTS	Direction of Trade Statistics
Dist	Great Circle Distance
EIU	Economist Intelligence Unit Riskwire & Democracy Index
E	Mega-Event Variable
FIFA WC	FIFA World Cup
F.O.B	Freight on Board
γ	Gama, Destination Fixed Effects
GCS	World Economic Forum Global Competitiveness Report
GDP	Gross Domestic Products
GDP PC	Gross Domestic Products Per Capita
GWP	Gallup World Poll
Govt Effec	Government Effectiveness
HER	Heritage Foundation Index of Economic Freedom
i	Country Destination
IFD	IFAD Rural Sector Performance Assessments
IJT	iJET Country Security Risk Ratings
IMF	International Monetary Fund
IPD	Institutional Profiles Database
j	Country Origin

λ	Lambda, Year Fixed Effects
Lang	Common Language Variable
LCU	Local Currency Units
Lions	Rugby Lions Tours
Ln	Natural log
OLS	Ordinary Least Square
ORK	Overall Risk Factor
Politcl Stab	Political Stability and Terrorism
POP	Population
PPP	Purchasing Power Parity
PRS	Political Risk Services International Country Risk Guide
Rgltry Qlty	Regulatory Quality
Rul of Law	Rule of Law
RWC	Rugby World Cup
SOG	Summer Olympic Games
t	Time, Years
Tou	Number of Tourists
TPR	US State Department Trafficking in People Report
u	Error Term
UNWTO	United Nations World Tourism Organization
WDI	World Development Indicators
WGI	World Governance Indicators
WMO	Global Insight Business Conditions and Risk Indicators
WOG	Winter Olympic Games
WRI	World Risk Indicators

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

PURPOSE AND PROBLEM STATEMENT

Mega sports events have been a research focal point for many years now. The enormous amount of investment on infrastructure development and marketing both from public and private sectors in the tourism industry have attracted many scholars to look into different aspects of hosting such events (Peng, Song, & Crouch, 2014). Despite the divide among scholars on the final outcome, the events are highly attended and competitions between countries for hosting them are ever rising. Fourie and Sanatno-Gallego, in 2011, introduced a Gravity Model to analyze the impact of mega events on inbound tourism over the 1995-2006 time period. In this thesis, more is built on their work and the analysis is extended for an additional seven years from 1995 to 2013. The model is also extended and new factors that largely influence tourist's intentions towards attending mega events are introduced.

Roche (2000) defines mega events as 'large-scale cultural (including commercial and sporting) events, which have a dramatic character, mass popular appeal and international significance' (p. 1). This definition is known as the best understandable way of describing the meaning of the term "Mega Events" (Hayes, & Karamichas, 2011). The definition highlights two highly important characteristics of mega events. First, the mega event shall have cultural significance on international levels, and second, it shall call for massive media coverage (Hayes et al. 2011). However, controversies over which event should be called a mega event and which not, does not end here. Some scholars (Hiller, 1995; Ritchie & Yangzhou, 1987; Rose & Speigel, 2011) study political summits, festivals and expos in the mega event context while others (Horne & Manzenreiter, 2006; Meannig & Zimbalist, 2012) consider some specific sports events as mega events. Muller (2014) further explores the mega events' concept and divides the events into three categories of 'Giga', 'Mega', and 'Major' based on a number of cost and benefit factors. In this thesis, the impact of six mega events namely, the Summer Olympic Games (SOG), the Winter Olympic Games (WOG), the FIFA World Cup (FIFA WC), the Cricket World Cup (CWC), the

Rugby World Cup (RWC) and the Rugby Lions Tours (Lions) is studied. It is assumed that, these events comply with Roche's (2000) definition of mega events as such that Fourie and Santano-Gallego (2011) included the same six types of mega sports events in their study.

Fourie and Santano-Gallego (2011), in their research, analyzed a set of predictors such as trade, GDP, and population among others and identified the significant ones. The same Gravity Model is applied to a data set with observations for an additional seven years. In this thesis, a three step approach is undertaken under three different sections towards the analysis. In Section I, two of the hypothesis from Fourie and Santano-Gallego's (2011) paper (sometimes referred to as the reference paper in this thesis) are tested over the same number of years and two different data sets, the first provided by Santano-Gallego and the second generated by author of this thesis. This is a compare-and-validate section where results from the two data sets are compared for validation purposes. In the second part, all four hypothesis in Fourie and Santano-Gallego's work (2011) are re-tested over the extended data set. New findings are captured and changes are discussed. In the third part of the analysis, the Gravity model is extended and new control variables are included in the three dimensional study. The impact of mega events on the international inbound tourism is revisited and findings are reported. The literature review of mega events, tourism, and their direct or indirect relationship with other significant factors presented next highlights important aspects of the study undertaken.

REVIEW OF THE LITERATURE

International tourism being one of the rapid growing industry has gained significant attention leading to extensive research on tourism demand forecasting (Claveria, Monte, & Torra, 2015). In the last two decades several tourism demand forecasting models have been proposed that could be grouped under different categories of Qualitative versus Quantitative, Simplistic and Stochastic, Linear versus Non-Linear, and the more distinguished category of Time Series versus Econometric models. The enormous amount of research in the area is persuasive in nature, leading scholars to perform reviews of the aforementioned modeling techniques. For example, Li, Song, and Witt in 2005 reviewed 84 post 1990 empirical tourism forecasting models, Crouch (1994) studied 85 empirical practices, Lim (1999) meta-analyzed 70 articles in search of a more generalized approach, and Song and Li in 2008 studied 121 papers published between 2000 and 2007. The reviews delineate substantial insight into the literature. As such, it is widely accepted that the

performance and accuracy of all the modeling techniques strongly depend on the sample size, and selection of the control variables. Moreover, in most of these reviews, the Econometric models are encountered more often, emphasizing on its popularity among scholars. For example, in Song and Li's (2008) review of the literature 71 of the 121 papers included uses Econometric models and 30 of them used a combination of Econometric and Time Series.

As mentioned above, the variety in selection of deterministic variables is an important aspect of international tourism demand forecasting models. Often, tourism is referred to as a form of global trade in services (Fourie & Santano-Gallego, 2011; Jensen & Zhnag, 2013). This understanding of tourism initiates strong ties between international tourism and trade. Numerous scholars in the Economics industry have studied the possible relationship and impact of one over another (Keum, 2008; Morley, Rosello, & Santana-Gall, 2014; Narayan & Naguyan, 2015; Santano-Gallego et al*.,* 2015). The research results show a number of scholars confirm that International tourism leads to economic growth. However, the number of data points included in the studies which largely influence accuracy of the results are relatively small. Price levels and relative prices are another significant factor that could be often seen in the tourism literature (Kim & Lee, 2016). Gross Domestic Products per capita (GDP PC) and population are the most common indicators in the Econometric studies. Common currency, common language, common border, and geographical distance have also been considered significant cultural and geographical decisive factors in determining international tourism (Akerlof, Rose, Yellen, & Hessenius, 1991; Eilat & Einav, 2004; Rose, 2000). Although these factors and their impact on tourism have been studied by different scholars individually or a combination of two or more, it is very rare to find models that incorporate all the significant variables, especially when it comes to the study of the relationship between mega sports events and international tourism.

Researchers have looked at the concept of Mega Events from different perspectives. Some scholars have focused on the socio-political, environmental, and developmental consequences of these events (Barker, Page, & Meyer, 2002; Hiller, 1998; Szymanaski, 2001), while others have studied the publicity, image building and the lasting legacies of the host cities (Nyikana, Tichaawa, Swart, 2014; Kim & Chalip, 2003; Lee, Taylor, Lee, & Lee, 2005). Little attention has been paid to its impact on international tourism. One of the important papers on the topic is Rose and Speigel's (2011) work where they studied the impact of mega events on international trade. Fourie

and Santano-Gallego (2011) use the same methodology as Rose and Spiegel's (2011) and studied the impact of mega events on international tourism. The approach to the analysis in this thesis is inspired by Fourie and Santano-Gallego (2011). More is built on their work for an additional seven years, and the Gravity Model is extended, so, as suggested by many scholars, that combination of more predictors will yield higher accuracy in forecasting than using single prediction (Andrawis, Atiya, & El-Shishiny, 2011; Costantini & Pappalardo, 2010).

CHAPTER II

DATA AND METHODOLOGY

THE GRAVITY MODEL

Researchers used what is known as the Spatial (Gravity) model first introduced by Tinbergen in 1962 to predict the number of tourists traveling between pairs of countries and to identify which predictors are more significant than others. The basic concept is based on Newton's Law of Universal Gravitation, in which the gravitational force between two objects is directly proportional to their masses and inversely proportional to the squared distance between them. The idea was adapted for trade and tourism and was developed in the 60s and 70s using the same formula of $F_{ij}=g\ m_1\ m_2/d_{ij}^2$ where F_{ij} represents in this case the trade flow between two countries i and j; m_i and m_j are their economic sizes; d_{ij} is the distance between them; and g is a constant. This relation means that trade flows between two countries are proportional to the scale of their economies and inversely affected by the distance between them (Keum, 2010). Since then, the model has gone through several iterations of development by several researchers to predict the amount of trade, and then the number of international tourists as a form of trade commodity, and to also identify the significant predictors. Fourie and Santana-Gallego (2011) presented the following Gravity Model:

$$\begin{aligned} Ln\ Tou_{ijt} = {} & \beta_0 + \beta_1 LnTrade_{ijt} + \beta_2 LnGDPpc_{it} + \beta_3 LnGDPpc_{jt} + \beta_4 LnPOP_{it} \\ & + \beta_5 LnPOP_{jt} + \beta_6 LnPPP_{ijt} + \beta_7 LnDist_{ij} + \beta_8 Lang_{ij} \\ & + \beta_9 Border_{ij} + \beta_{10} Colony_{ij} + \beta_{11} CU_{ij} + \acute{\eta} E_{it} + \gamma_i + \delta_j + \lambda_t \\ & + u_{ijt} \dots\dots\dots\dots \end{aligned} \quad (1)$$

where

Ln: Natural log

i: Destination country

j: Origin country

Tou_{ijt}: Number of tourists between *i* and *j* at year *t*

$Trade_{ijt}$: Real bilateral trade-in-goods, as the sum of exports and imports, between *i* and *j*

$GDPpc_{it}$: GDP per capita of *i* in year *t*

$GDPpc_{jt}$: GDP per capita of j in year t

POP_{it}: Population of i in year t

POP_{jt}: Population of j in year t

PPP_{ijt}: Purchasing power parity that reflects relative cost of living in the i with respect to j

$Dist_{ij}$: Great circle distance between the capital cities of i and j

$Lang_{ij}$: 1 if there is a common language between i and j; 0 otherwise

$Border_{ij}$: 1 if there is common land border between i and j; 0 otherwise

$Colony_{ij}$: 1 if there has ever existed colonial relationship between i and j; 0 otherwise

CU_{ij}: 1 if I and j share common currency; 0 otherwise

E_{it}: 1 if a mega-event is held at i in year t; 0 otherwise

γ_i: Destination fixed effect

δ_j: Origin fixed effect

λ_t: Year fixed effect

u_{ijt}: Error

The authors in the reference paper used a data set that includes 169 countries as tourist destination and 200 countries as origin of tourists over the period 1995 – 2006 (33,800 pairs of countries). They obtained the number of annual international tourist arrivals by country of origin from the United Nations World Tourism Organization (UNWTO). The sources of their input data are listed in the reference paper. Ordinary Least Square (OLS) regression was used to study the significance of selected predictors. While the results of their analysis are discussed in detail in Section I and Section II, the same model is applied to exactly the same pairs of countries in the updated and extended analysis. Country fixed effects of origin and destination and year fixed effects are included in the OLS model. Heteroscedasticity (inconsistency in the variance of the standard error) in the data set is also taken into account, such that the robust standard error is clustered by country pairs. In the tables of results, coefficient estimates of the variables are the main determinants of the level of increase or decrease in tourism if all other variables are held constant. Moreover, significant coefficient estimates at Alpha (α) level of 0.99 are marked with double asterisks (**), those significant at α level of 0.90 are marked with a single asterisk (*), and estimates found to be statistically insignificant are left unmarked. The upper and lower confidence intervals for all the analysis results have been include in Appendix D.

DATA COLLECTION

A new data set for the input variables included in Equation 1 has been generated. The definitions, principles and criteria explained below constitute the foundations of the new data development for the extended analysis. Efforts have been made to not-to deviate from the original method of data set construction.

1. **Country Pairs:** In order to keep consistency with the original data set used by authors in the reference paper, the same pairs of countries are included in the new data set and hence in the extended analysis. Pairs of the countries are selected based on the fundamental theorem of each pair having considerable ties with the counterpart country over one or more independent variables. List of all countries used as tourism destinations and those used as tourism origins could be found in Appendix A (A1 and A2), respectively.
2. **Inbound Tourist Arrivals:** Data on tourism is obtained from the United Nations World Tourism Organization (UNWTO). The organization is recognized by the United Nations (UN) as an appropriate source for data collection, analysis, and publication (Haya, 2015). The UNWTO provides statistics for different types of international tourism, such as:
    - Inbound Tourism
    - Domestic Tourism
    - Outbound Tourism
    - Tourism Industries,

    Inbound tourism statistics from the list above is closest in nature to the purpose of this study, hence included in the dataset. According to the UNWTO, inbound tourism captures arrival of non-resident tourists visiting country of reference regardless of their travel purpose. Each visit is counted as a separate trip whether completed by the same person or a different one. The UNWTO's clear description of characteristics helps in reducing limitations on data sets encountered in the literature (Eilat & Einav, 2004). Although, the organization further breaks down the data into different categories, the number of non-resident visitors who have had an overnight-stay specifically is included in the analysis. The UNWTO refers to administrative records, immigration, traffic counts, border surveys, or a combination of one or more of the mentioned indicators from individual countries as source of data collection (Haya, 2015).

3. **Bilateral Trade Flow:** The source of data for this variable is the Direction of Trade Statistics (DOTS) database of International Monetary Fund (IMF). IMF's DOTS provides data for the value of merchandise imports-from and exports-to the most important trade partners of around 184 countries. This organization breaks down the data set into different categories from which the country by partner data is used for analysis in this thesis (DOTSY, 2015). According to *A Guide to Direction of Trade (1993),* the method for data collection from the member countries is based on the custom and border documents and the foreign exchange control record. The organization encourages member countries for their contribution of data submission by giving them flexibility in reporting currency. Countries can submit their trade data in U.S. dollars or their local currency for the merchandise exports to the destination (F.O.B) and imports from origin (C.I.F). During the process of trade data collection for use in this thesis, some differences were spotted for the total trade values between partnering countries. For instance: When considering a pair of countries, one as a destination and the other as the origin, the total trade (sum of exports and imports) would have different values from those if their assignments as destination and origin were switched. The reasons for these changes which in some cases may cause questioning validity of the data set are explained by IMF at www.imf.org as:
    a) Some of the countries fail to report the data on regular basis, and therefore DOTS uses statistical methods to estimate the value of the trade based on previously reported data of the country of concern or its trade partnering country.
    b) Sometimes the source country reports the total value of the exports to the destination country in their reports, however the destination country is used as a transit to a third country and does not include them in their import data.
    c) And, finally, in some cases, the shipment time makes the same exports from the origin country appear in another year's trade data for the destination country, due to the merchandise transit and shipment time which is directly proportional to distance between the two countries. These differences are small and do not have a major impact on the overall analysis. The data entries for the purpose of analysis in this thesis are in thousands of dollars ($US) and are converted into real terms with the use of U.S. GDP Deflator, of base year 2010.
4. **GDP Per Capita and GDP:** GDP per capita (GDP PC) is another economic indicator included as an independent variable in Equation 1. GDP PC and bilateral trade are the most

common factors in most of the published research work concerning tourism forecasting and analysis. Data for GDP per capita was obtained from the World Development Indicators (WDI) data set of the World Bank. WDI contains more than 1300 time series cross-country developmental indicators for around 214 economies (WDI, 2015), however the indicators of interest for the analysis purpose in this study are GDP per capita, GDP, Population and purchasing power parity (PPP) for specific countries. GDP per capita is converted into real terms using two of the WDI indicators:

a) GDP per capita, PPP (constant 2011 international $).
b) PPP conversion factor from U.S. Dollars to International Dollars.

The following formula is used to calculate GDP Per capita:

$$\text{GDP PC} = \left(\text{GDP per capita, PPP (constant 2011 international \$)}\right) * \frac{1}{(\text{PPP Factor from International Dollars to U. S. dollars})} \dots\dots\dots\dots \quad (2)$$

This method prepares the GDP PC values to be on the same scale as bilateral trade. GDP values in the WDI are reported in current U.S. dollars. Therefore, to convert them into real terms, the U.S. GDP deflator is used applying the following formula:

$$\text{GDP} = \frac{\text{GDP at market prices (current US\$)}}{\text{US GDP Deflator Base2010}} \dots\dots\dots\dots\dots \quad (3)$$

In this way the data is consistent and in the same format.

5. **Purchasing Power Parity (PPP):** PPP is another important economic indicator for comparison between countries. This variable too has extensively been used by scholars to develop an origin-destination relationship. The World Bank defines PPP as the number of local currency units of a country required to buy the same amount of goods and services in the market as the U.S. dollars would buy. Country-specific PPP values are obtained by taking the ratio of GDP reported in the current Local Currency Units (LCU) and GDP-PPP reported in current international dollars in the WDI data set. As a result, the PPP would represent the relative difference of country and the United States (used as a benchmark for all countries at the first place), since the interest of this study lies in the relative prices of country pairs with each other; therefore, the ratio of the PPPs of country pairs is included in the extended data set.

6. **Population:** Total population for each country is included in the analysis as provided in the WDI data set.
7. **Mega Events:** Fourie and Santano-Gallego (2011)studied the impact of hosting six Mega Events on tourist arrival. Namely, the Summer Olympic Games (SOG), the Winter Olympic Games (WOG), the FIFA World Cup (FIFA WC), the Cricket World Cup (CWC), the Rugby World Cup (RWC), and the Lions Tours (Lions). In order to be able to validate their results and thereafter build on them for an extended time, the same six type of mega events hosted over the years 1995-2013 are included in the study. Necessary information about the time and place of the events, bid candidates and host countries, and participating countries in the mega events have been collected from their respective official website. For example, information on the SOG and the WOG was obtained from www.olympic.org, the FIFA WC from www.fifa.com/worldcup, the RWC from www.rugbyworldcup.com, the CWC from www.icc-cricket.com/cricket-world-cup, and finally data for Rugby Lions tours were obtained from www.lionsrugby.com.
8. **Common Language:** This binary variable captures the effects of common spoken language and its impact on international tourism and data were obtained from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) database. The CEPII database for language proximity criterion requires a language to be spoken by at least 9-20% of the total population of both countries and be one of the up to three official languages to be considered common. CEPII refers to the web site www.ethnologue.org and the CIA World Fact Book as original source for the data collection. According to Melitz and Toubal (2014), who studied the impact of common language on bilateral trade, it is important for businesses to invest in learning the language of their trade partners in order to effectively communicate trade related deficiencies and issues.
9. **Contiguity and Distance:** Contiguity is another variable that is often used by economists to develop correlation for bilateral trade between countries. Data for this variable were obtained from the CEPII data-base. The CEPII Geodesic distance uses geographical coordinates (Latitude and Longitude) and the "Great Circle" method of calculation to calculate distance between main cities of the country pairs under study. Anderson and Wicoop (2001) uses the same method for similar distance calculations. In the CEPII database for 13 of the total 225

countries, the capitals were not populated enough to be considered as the main city, therefore a more populated city is considered as the economic center.

10. **Colony:** This dummy (binary) variable included in Equation 1 concerns colonial ties between countries. The CEPII database incorporates three different types of colonial relationships between countries in their data set.
    a) Common Colonizer, which lists all pairs of countries that have been colonized by the same country after 1945,
    b) Currently in Colonial Relationship, which lists the current colonizer and colonized pairs of countries, and
    c) Colony, which lists pairs of countries that has ever had any colonial links.

    The third indicator is more comprehensive and aligns well with the purpose of this study, hence is included in the data set.

CHAPTER III

ANALYSIS RESULTS AND DISCUSION

In Section I of this chapter, two of the four predominant hypothesis presented in Tables 1 and 3 of Fourie and Santano-Gallego's (2011) work are re-tested and the original results are replicated. In this same section, the same methodology is applied to the updated data set over the years 1995-2006 and the results are compared. The objective in doing so is validation of the model, methodology and the data update process. In Section II, the validated method and model in Section I is applied to the updated data set over the years 1995-2013. In this part of the study, the changes in the results, consequent to additional observations are captured and discussed. Section III is allocated to the extension of the Gravity Model, data set, and discussion of the results of the extended analysis.

SECTION I: VALIDATION OF THE MODEL AND METHOD

As mentioned above, first the Ordinary Least Square (OLS) regression modeling technique is applied to the data set provided by Fourie and Santano-Gallego (2011), using Equation 1 and the "R" statistical package in order to replicate the original results presented in Tables 1 and 3 of their paper. Next, the same methods and techniques are applied to the updated data set for years 1995-2006 and the results are compared. In this way the basics of the extended analysis are explained, and foundations are laid.

Impact of Mega Events and Tourist Arrivals (Hypothesis 1)

The concept of whether hosting mega sports events increase tourism is the founding hypothesis and of foremost importance in the reference paper. It constitutes as the center-of-mass for the rest of the three hypothesis in Fourie and Santano-Gallego's (2011) work. Table 1 presents coefficient estimates and statistical significance of the variables using the original dataset.

Table 1

Impact of Mega Events on Tourism– Original Dataset

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	11.8282975 **	3.9140	9.086E-05	11.7385386 **	3.8853	0.0001023
Ln Trade$_{ij}$	0.0691017 **	19.6676	< 2.2E-16	0.0690982 **	19.6661	< 2.2E-16
Ln GDPPC$_j$	0.2686685 **	8.2000	2.438E-16	0.2685700 **	8.1995	2.448E-16
Ln GDPPC$_i$	0.1703710 **	5.3061	1.123E-07	0.1711231 **	5.3236	1.020E-07
Ln POP$_j$	0.0228584	0.1328	0.8943706	0.0274706	0.1596	0.8732005
Ln POP$_i$	-0.0746186 *	-2.1997	0.0278291	-0.0741453 *	-2.1881	0.0286678
Ln PPP$_{ij}$	-0.0322355 **	-2.6536	0.0079646	-0.0326063 **	-2.6835	0.0072881
Ln Dist$_{ij}$	-1.4822426 **	-63.9514	< 2.2E-16	-1.4822567 **	-63.9499	< 2.2E-16
Lang$_{ij}$	1.0753988 **	21.7327	< 2.2E-16	1.0754257 **	21.7326	< 2.2E-16
Border$_{ij}$	1.1839538 **	10.6566	< 2.2E-16	1.1838619 **	10.6553	< 2.2E-16
Colony$_{ij}$	0.9181783 **	8.0237	1.039E-15	0.9181887 **	8.0235	1.041E-15
CU$_{ij}$	0.2288761*	1.8927	0.0583988	0.2286064	1.8903	0.0587175
Event	0.0778742 **	6.2834	3.328E-10			
SOG				0.1776456 **	5.2312	1.689E-07
WOG				-0.0684410 **	-3.3400	0.0008381
FIFA				0.0764692 **	2.7835	0.0053784
CWC				0.1928068 **	5.7081	1.146E-08
RWC				-0.1232154 **	-3.1886	0.0014301
Lion				0.1461575 **	4.4232	9.740E-06
Observation	83,520			83,520		
F-statistics	213.63		0.00	211.06		0.00
R-Squared	0.8376			0.8376		

The overall model based on 83,520 observations with a 0.8376 R-Squared value is significant enough. The T-Values in Table 1 shows that, all the variables included in the model are statistically significant except for the population of origin. However, the level of significance is lower for the common currency and population of destination variables. Next is Table 2 for comparison purposes. The same OLS regression analysis when applied to the updated data set using Equation 1, and observations over the years 1995-2006 are included.

Table 2

Impact of Mega Events on Tourism– Results of Extended Dataset

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	6.7712512	1.9398	0.0524054	6.5580301	1.8782	0.0603588
Ln Trade$_{ij}$	0.3014933 **	32.3973	< 2.2E-16	0.3015367 **	32.4011	< 2.2E-16
Ln GDPPC$_j$	0.1072690 **	3.0790	0.0020777	0.1071137 **	3.0762	0.0020973
Ln GDPPC$_i$	0.3053537 **	9.7050	< 2.2E-16	0.3064180 **	9.7275	< 2.2E-16
Ln POP$_j$	-0.2889395	-1.8703	0.0614458	-0.2825307	-1.8297	0.0672996
Ln POP$_i$	0.2185972	1.3850	0.1660668	0.2252177	1.4263	0.1537796
Ln PPP$_{ij}$	-0.0522931 **	-3.1540	0.0016110	-0.0532859 **	-3.2132	0.0013131
Ln Dist$_{ij}$	-1.0825052 **	-42.0469	< 2.2E-16	-1.0824408 **	-42.0438	< 2.2E-16
Lang$_{ij}$	0.8164163 **	17.1009	< 2.2E-16	0.8164394 **	17.1009	< 2.2E-16
Border$_{ij}$	1.1027455 **	10.7744	< 2.2E-16	1.1026519 **	10.7730	< 2.2E-16
Colony$_{ij}$	0.6786857 **	5.9897	2.113E-09	0.6787652 **	5.9907	2.099E-09
CU$_{ij}$	0.1925043	1.8072	0.0707304	0.1923536	1.8058	0.0709554
Event	0.0680087 **	5.3660	8.074E-08			
SOG				0.2009416 **	5.6468	1.641E-08
WOG				-0.0774044 **	-3.3115	0.0009283
FIFA				0.1187005 **	4.5085	6.539E-06
CWC				0.2232360 **	6.8098	9.850E-12
RWC				-0.1424250 **	-3.6759	0.0002372
Lion				0.0335407	0.9625	0.3357918
Observation	72,213			72,213		
F-statistics	208.12		0.00	205.61		0.00
R-Squared	0.8389			0.8389		

When comparing the results of Table 2 and Table 1, most of the variables maintain the same statistical significance. Exceptions are the intercept, population of the destination country, and common currency. In contradiction to the results over the updated data set, intercept is statistically

very significant in Fourie and Santano-Gallego's (2011) paper. The population of the destination country and common currency were statistically significant at (α =0.90) significance level in the original 2011 analysis and are statistically insignificant now. Although most of the coefficients estimates remain nearly unchanged, there are a few changes in the value estimate and sign. The sign for the population of destination's variable has changed from negative to positive and vice versa for population of origin variables. Furthermore, the coefficient estimates for trade and population have higher values than Table 1. Although some of these changes seem more explicable in Table 2 than in Table 1, a conclusion is not drawn here, as the purpose of the analysis in the current section is validation of the updated data set. Therefore, the focus is made on the reasons behind these changes, and based on the knowledge of the data set developmental process, there are two possibilities for that:

a. First, the values of trade and GDP PC are included in the updated data set on a different scale with a U.S. GDP Deflator and base year 2010 than in original data set (base year 2000). This can have an impact on differences in the coefficient estimates as well as signs and statistical significance.
b. Secondly, the number of observations for the years 1995-2006 in the updated data set (72,213) are less than the number of observation in the original data set (83,520) which can also have an impact on the overall model.

Despite the presence of minor alterations, most of the coefficients estimates, signs, and statistical significance of the variables remain very similar and confirm on the validation of data set generation process.

Participating partners and seasonal effects (Hypothesis 3)

In another attempt, the third hypothesis of whether tourism for the countries participating in the games are more than those not participating, and whether a mega sports event held in the peak tourism season versus one held in the off-peak tourism season has different results are tested. Fourie and Santan-Gallego (2011) defines summer as the peak season for tourism and defines fall, winter and spring as off-peak tourism seasons. Therefore, the binary variable Event Peak takes the value of one (1) if the event is hosted in the summer, zero (0) otherwise. The same is true for the binary variable off peak, it takes the value of one (1) if the event was hosted in fall, winter, or

spring, and zero (0) for summer. The event participation binary variable takes the value one (1) if the team of the origin-country for the specific pair of countries under study has played in the games, zero (0) otherwise. The non-participant variable is the opposite of the participant binary variable.

Table 3

Participating Partners and Seasonal Effects of Mega Events – Results of Original Dataset

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	11.7735454 **	3.8952	9.819E-05	11.8964508 **	3.9352	8.321E-05
Ln $Trade_{ij}$	0.0691129 **	19.6691	< 2.2E-16	0.0691070 **	19.6703	< 2.2E-16
Ln $GDPPC_j$	0.2688091 **	8.2042	2.353E-16	0.2684863 **	8.1942	2.558E-16
Ln $GDPPC_i$	0.1703034 **	5.3016	1.151E-07	0.1706870 **	5.3166	1.060E-07
Ln POP_j	0.0262965	0.1527	0.8786128	0.0185727	0.1078	0.9141194
Ln POP_i	-0.0749006 *	-2.2046	0.0274883	-0.0743958 *	-2.1957	0.0281146
Ln PPP_{ij}	-0.0324298 **	-2.6689	0.0076112	-0.0318944 **	-2.6248	0.0086709
Ln $Dist_{ij}$	-1.4822475 **	-63.9498	< 2.2E-16	-1.4821948 **	-63.9473	< 2.2E-16
$Lang_{ij}$	1.0752162 **	21.7290	< 2.2E-16	1.0754013 **	21.7325	< 2.2E-16
$Border_{ij}$	1.1838308 **	10.6550	< 2.2E-16	1.1839547 **	10.6567	< 2.2E-16
$Colony_{ij}$	0.9178556 **	8.0197	1.073E-15	0.9182013 **	8.0239	1.038E-15
CU_{ij}	0.2285182	1.8891	0.0588845	0.2300616	1.9030	0.0570377
Participant	0.1175945 **	3.9817	6.849E-05			
None Participant	0.0030567	0.1095	0.9128004			
Peak Season				-0.0302324	-1.1591	0.2464246
Off Season				0.1672313 **	7.5270	5.244E-14
Observations	83,520			83,520		
F-Statistics	213.11	0.00		213.01	0.00	
R-Squared	0.8376			0.8376		

Based on results as shown in Table 3, Fourie and Santano-Gallego (2011) recommend hosting a mega sports event in the off-peak season. They also suggest that countries participating in the mega events have more tourism gains while the countries not participating are not statistically significant.

The analysis has been performed on the updated data set and the results shows:

Table 4

Participating Partners and Seasonal Effects of Mega Events – Results of Extended Dataset

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	6.6664508	1.9092	0.0562447	6.8816708 *	1.9702	0.0488175
Ln Trade$_{ij}$	0.3015366 **	32.4007	< 2.2E-16	0.3014754 **	32.3955	< 2.2E-16
Ln GDPPC$_j$	0.1073734 **	3.0825	0.0020534	0.1072130 **	3.0772	0.0020902
Ln GDPPC$_i$	0.3051252 **	9.6964	< 2.2E-16	0.3051681 **	9.6954	< 2.2E-16
Ln POP$_j$	-0.2848963	-1.8442	0.0651602	-0.2914831	-1.8858	0.0593293
Ln POP$_i$	0.2211621	1.4012	0.1611646	0.2141535	1.3564	0.1749841
Ln PPP$_{ij}$	-0.0527053 **	-3.1795	0.0014758	-0.0520833 **	-3.1390	0.0016960
Ln Dist$_{ij}$	-1.0824728 **	-42.0452	< 2.2E-16	-1.0825328 **	-42.0473	< 2.2E-16
Lang$_{ij}$	0.8162815 **	17.0979	< 2.2E-16	0.8164245 **	17.1011	< 2.2E-16
Border$_{ij}$	1.1025889 **	10.7727	< 2.2E-16	1.1027878 **	10.7748	< 2.2E-16
Colony$_{ij}$	0.6783038 **	5.9858	2.164E-09	0.6786840 **	5.9894	2.116E-09
CU$_{ij}$	0.1924133	1.8062	0.0708839	0.1925617	1.8078	0.0706448
Participant	0.1425823 **	4.7276	2.276E-06			
Non Participant	0.0698092 *	2.1901	0.0285179			
Peak Season				0.0107327	0.4174	0.6763932
Off Season				0.1018887 **	4.2999	1.711E-05
Observations	72,212			72,212		
F-Statistics	207.54		0.00	207.5		0.00
R-Squared	0.8389			0.8389		

The coefficients' estimates over the updated analysis remain close to those in Fourie and Santano-Gallego's (2011) paper, both for the seasonality and participation variables. The statistical significance of the variables also remains consistent for peak season, off-peak season, and event participation. The only change occurs in statistical significance of the non-participant variable, which slightly increases to α=0.90 significance level. Again, no conclusion is drawn here and the marginal increase is assumed to be associated with the two datasets and time-linked reasons mentioned above.

Hypotheses Two and Four were also tested and there were no noticeable changes. Thus, it is assumed that despite the minor changes in the values and statistical significance of the coefficients for some variables, over all the data set replicates the same results and therefore is valid for further analysis.

SECTION II: EXTENDED TIME FRAME ANALYSIS

In this section, the OLS model is applied to the updated data set using Equation 1 and approximately 50% additional observations to capture changes in the behavior of the independent variables. The coefficients of the variables in Equation 1 are main determinants of the level of increase or decrease in tourist arrivals if all other indicators are held constant. The statistical analysis is performed using "R" statistical package. The advantage of using R is its capability of clustered analysis. It allows for clustering the robust standard error, which in turns accounts for the heteroscedasticity in the data set.

Impact of Mega Events on Tourism

First of all, the hypothesis of whether hosting a mega sporting event, as defined in Chapter II, increases tourism between countries is tested. Both the individual and overall impact of all six mega sports events are studied and results of the analysis are presented in Table 5. Before the impact of the mega events is discussed, it is necessary to confer the overall model and the deterministic control variables. Table 5 shows that the extension of the data set increases R-squared value of the model by almost 1% to 84%. Having around 40,000 additional observations, indeed, is a good improvement towards better results. It also confirms on the fact that bilateral trade, GDP PC of origin, and GDP PC of destination are significantly positive and are of greater importance in determining tourism between countries. This is also a confirmation on the Fourie and Santano-Gallego's (2011) argument that the richer the countries the higher the tourism flow.

Although, the population of the tourism origin country remains statistically insignificant, the population of the destination country is significantly positive. One of the important results of the model with the extended data is the change in sign and significance of population destination from negative (-0.0746) to positive (0.1916) when compared to Foruie and Santano-Gallego's (2011) work in the reference paper, that argues that the inclusion of GDP PC in the model accounts for the demand size and therefore the population of the destination country is not important. However, studies suggest that GDP PC cannot be deterministic of population size neither its economic importance (Birchenall, 2016; Singha & Jaman, 2013). Therefore, the results of the analysis could be explained by presumptive direct relationship between population and publicity, economic growth, and technological advancements which in turns can indirectly promote tourism.

Table 5

Impact of Mega Sports Events on Tourist Arrival

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	2.7452128	1.4708	0.1413586	2.6899000	1.4408	0.1496537
Ln Trade$_{ij}$	0.2855140 **	32.9508	< 2.2E-16	0.2855200 **	32.9516	< 2.2E-16
Ln GDPPC$_j$	0.1849429 **	6.9133	4.759E-12	0.1848200 **	6.9124	4.789E-12
Ln GDPPC$_i$	0.3028314 **	12.8928	< 2.2E-16	0.3017500 **	12.8362	< 2.2e-16
Ln POP$_j$	-0.0054206	-0.0634	0.9494759	-0.0040787	-0.0477	0.9619519
Ln POP$_i$	0.1916328 *	2.2294	0.0257878	0.1944600 *	2.2618	0.0237126
Ln PPP$_{ij}$	-0.1072161 **	-7.7201	1.171E-14	-0.1078600 **	-7.7694	7.948E-15
Ln Dist$_{ij}$	-1.0838705 **	-43.9893	< 2.2E-16	-1.0839000 **	-43.9878	< 2.2E-16
Lang$_{ij}$	0.8322578 **	18.5158	< 2.2E-16	0.8322400 **	18.5152	< 2.2E-16
Border$_{ij}$	1.1160010 **	11.4539	< 2.2E-16	1.1160000 **	11.4532	< 2.2E-16
Colony$_{ij}$	0.5963257 **	5.3952	6.856E-08	0.5963800 **	5.3959	6.832E-08
CU$_{ij}$	0.2119878 *	1.9999	0.0455171	0.2119100 *	1.9991	0.0455964
Event	0.0340307 **	3.1796	0.0014753			
SOG				0.1996300 **	7.4202	1.177E-13
WOG				-0.1058100 **	-5.6901	1.272E-08
FIFA				0.0801770 *	2.3116	0.0208016
CWC				0.1680700 **	6.5307	6.570E-11
RWC				-0.1192800 **	-3.8988	9.673E-05
Lion				-0.0307830	-1.4582	0.1447892
Observation	122,747			122,747		
F-statistics	237.54		0.00	234.77		0.00
R-Squared	0.8411			0.8411		

Furthermore, the PPP and distance control variables are statistically significant with negative signs, meaning people intend to visit countries closer to them with lower differences in relative prices. The four dummy variables of common border, common language, currency union, and colonial ties are statistically significant with positive signs, indicating increase in tourism if countries under consideration hold a true value for one or more of these variables.

These results lay strong foundations for analysis henceforth. Yet, greater interest of this study lies in the impact analysis of mega sports events. Column one of Table 5 shows that the coefficient of hosting a mega sports event is significantly positive and confirms that mega events do increase tourist arrivals in the year the event is held. The disaggregated analysis of the six mega events however shows that the Rugby Lions tour is not significant at all in the tourism context. Low

density of the event's popularity on international scale can be one of the influential factors among others leading to this result. SOG, FIFA WC, and CWC are positively significant with varying level of increase in tourist arrivals such that SOG holds the highest, and FIFA WC the lowest positions. WOG and RWC are statistically significant with negative signs indicating demotion in tourism gains. Although this result complies with the findings in the reference paper and Rose and Spiegel's (2011) findings, the extended data set provides better foundations for explanation.

Table 6

Host Country by Mega Sports Event Type

Year	SOG	WOG	FIFA WC	CWC	RWC	Lions
1996	United States			India		
1998		Japan	France			
1999				United Kingdom	United Kingdom	
2000	Australia					
2001						Australia
2002		United States	Japan			
2003				South Africa	Australia	
2004	Greece					
2005						New Zealand
2006		Italy	Germany			
2007					France	
2008	China					
2009						South Africa
2010		Canada	South Africa			
2011				India	New Zealand	
2012	United Kingdom					
2013						Australia

In order to better understand the cognitive science behind the negative signs of Winter Olympic Games and the Rugby World Cup variables, the events in Table 6 are all listed in a synchronous format. The table shows that three out of the total four RWCs have been hosted in the same year

with a CWC, and all the four Winter Olympic Games have coincided with FIFA World Cups. Although there is possibility for several factors causing displacement in tourism, for instance, popularity density of specific mega sports events, number of countries participating in the event and location of the event could be playing important role in determining the negative signs. Yet, based on the evidence presented in Table 6, the coincidence of these events with one other in the same year could be outlined as the root cause for tourism displacement and a statistically demoting effect as a result. Based on the results of Tables 5 and 6, it would be reasonable to suggest the mega sports events' organizers to coordinate location and time of the two or more events coinciding with each other for better results.

The Lasting Legacy Effects of Mega Sports Events

The event's lasting legacy is often referred to as the most important benefit of hosting such events (Fourie & Santano-Gallego, 2011). In this part of the thesis, the events' lasting legacy hypothesis and their contribution to tourism gains is analyzed. Only the three mega events of SOG, FIFA WC, and CWC with highly positive economic significance have been studied over the immediate three years before and after the events.

Table 7

The Pre and Post-Mega-Events Impacts on Tourism

	Estimate	T-Value	Pr(>|t|)
(Intercept)	2.6931901	1.4415	0.1494407
Ln Trade$_{ij}$	0.2855295 **	32.9470	< 2.2E-16
Ln GDPPC$_j$	0.1844371 **	6.8943	5.438E-12
Ln GDPPC$_i$	0.3068537 **	13.0134	< 2.2E-16
Ln POP$_j$	-0.0037132	-0.0434	0.9653788
Ln POP$_i$	0.1911424 *	2.2227	0.0262379
Ln PPP$_{ij}$	-0.1083124 **	-7.7927	6.610E-15
Ln Dist$_{ij}$	-1.0838155 **	-43.9839	< 2.2E-16
Lang$_{ij}$	0.8322928 **	18.5167	< 2.2E-16
Border$_{ij}$	1.1161147 **	11.4547	< 2.2E-16
Colony$_{ij}$	0.5963463 **	5.3954	6.849E-08
CU$_{ij}$	0.2122886 *	2.0025	0.0452335
Event	0.0273481 *	2.0540	0.0399755
Event (t+1)	-0.0608087 **	-4.3056	1.667E-05
Event (t+2)	-0.0789803 **	-5.1909	2.096E-07
Event (t+3)	-0.0636052 **	-3.7864	0.0001529
Event (t-1)	0.0410887 * *	2.7441	0.0060679
Event (t-2)	0.0248628	1.8550	0.0636005
Event (t-3)	0.0288033	1.8955	0.0580228
Observations	122,747		
F-Statistics	234.16		0.00
R-Squared	0.8411		

Although it is widely claimed that hosting a mega sport events results in long term tourism gains, and so is confirmed by Fourie and Santano-Gallego (2011), the empirical results of the analysis in Table 7 contrarily suggest that positive gains from such events should be expected one year immediately before the event and in the same year of the event. The results indicate that the years after the event are statistically significant, however in the opposite direction. While this is unexpected and most of the countries struggle to host an event for long term gains, some of the qualitative factors causing this change could be the diversification behavior of tourists towards exploring new attractions. Visitors' level of satisfaction from the tour, costs, crowd management, and utilization of infrastructure of the host country can also have an impact on the attitude of the

tourists for a post event revisit. Most importantly, the pre-event marketing and publicity stops right after the event. Nevertheless, this opens a dialogue for further research in identification of areas to be improved in order to secure the expected lasting gains. Also, the fact that the facilities built for these events in the host cities become significant landmarks, and the event will leave enduring social, political, and developmental effects on the locals cannot be denied (Gursoy & Kendall, 2006; Shi, Yu, & Chen, 2015). The hypothesis whether seasonality and participation in mega events have any impact on the overall gains from such events are analyzed next.

Participation and Seasonality Impact of Mega Events

Results of the analysis presented in Table 8 validates the findings in the reference paper that suggested high tourism gains from the countries participating in the sport event while the none participating countries remain insignificant. The increase in tourism could range from 17.78% to almost 18%. Fourie and Santano-Gallego (2011) suggest that hosting mega events are significant only in the off peak season. Contradictory to their results our findings shows that hosting the event both in-peak and off-peak tourism seasons are significant.

Table 8

Event Participation and Seasonality Results

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	2.6918071	1.4423	0.1492298	2.72348148	1.4598	0.1443556
Ln Trade$_{ij}$	0.2855485 **	32.9562	< 2.2E-16	0.28549068 **	32.9483	< 2.2E-16
Ln GDPPC$_j$	0.1850713 **	6.9202	4.533E-12	0.18551430 **	6.9375	4.010E-12
Ln GDPPC$_i$	0.3020322 **	12.8586	< 2.2E-16	0.30304441 **	12.9099	< 2.2E-16
Ln POP$_j$	-0.0033858	-0.0396	0.9684117	-0.00290827	-0.0340	0.9728654
Ln POP$_i$	0.1933245 *	2.2493	0.0244979	0.18988921 *	2.2101	0.0271006
Ln PPP$_{ij}$	-0.1076146 **	-7.7517	9.139E-15	-0.10721499 **	-7.7191	1.180E-14
Ln Dist$_{ij}$	-1.0837687 **	-43.9850	< 2.2E-16	-1.08389764 **	-43.9899	< 2.2E-16
Lang$_{ij}$	0.8320185 **	18.5125	< 2.2E-16	0.83222148 **	18.5150	< 2.2E-16
Border$_{ij}$	1.1157900 **	11.4520	< 2.2E-16	1.11606898 **	11.4547	< 2.2E-16
Colony$_{ij}$	0.5956912 **	5.3895	7.077E-08	0.59630340 **	5.3952	6.858E-08
CU$_{ij}$	0.2119600 *	1.9997	0.0455377	0.21208236 *	2.0007	0.0454268
Participant	0.1636817 **	6.9973	2.623E-12			
Non Participant	-0.0380526	-1.6856	0.0918749			
Peak Season				-0.09491010 **	-5.7654	8.166E-09
Off Season				0.10990090 **	5.9696	2.385E-09
Observations	122,747			122,747		
F-Statistics	237.26		0.00	236.99		0.00
R-Squared	0.8411			0.8411		

The analysis results with the negative sign for hosting the mega event in peak tourism season (defined as summer) strongly suggest to avoid such times. Although the authors of the reference paper have similar suggestion based on the statistical significance of the Tourism off-peak season (Fall, Winter, and Spring), the statistical significance of the tourism peak season with a negative sign in this thesis using the extended data further strengthens the argument. This means that while there will be tourists visiting the country regardless of the event in the peak season, the event's impact will be less than expected, and it will reduce the overall impact of tourists' attraction by mega sporting event by almost 4%. Therefore, it is suggested for event's organizers to plan mega events in off-peak seasons in order to achieve the targeted gains. Another important factor in Table 8 is the event participation variable. From the results, it is concluded that a positive impact of mega sport event is associated with the countries who directly participate in the events.

Participation in the Bidding Process and Its Impact on Inbound Tourism

Rose and Spiegel (2011) suggest an almost equal consequential effects for countries that participate in the bidding process with the countries that actually host the mega events, and Fourie and Sanatano-Gallego (2011) confirm on their results. In contradiction to their findings, the results of analysis over the extended time frame presented in Table 9 suggest that participation in the bidding process will not have significant impact on increasing or decreasing tourism. Moreover, there is not enough evidence in the literature to support the argument of participation in the bidding process leading to economic benefits from this industry.

Table 9

Participation in Bidding for Hosting Mega Sports Event and Its Impact on Tourism

	Estimate	T-Value	Pr(>|t|)
(Intercept)	2.7287061	1.4621	0.1437164
Ln Trade$_{ij}$	0.2855207 **	32.9518	< 2.2E-16
Ln GDPPC$_j$	0.1847954 **	6.9076	4.952E-12
Ln GDPPC$_i$	0.3031578 **	12.9021	< 2.2E-16
Ln POP$_j$	-0.0056987	-0.0666	0.9469033
Ln POP$_i$	0.1929525 *	2.2445	0.0248020
Ln PPP$_{ij}$	-0.1072772 **	-7.7248	1.128E-14
Ln Dist$_{ij}$	-1.0838656 **	-43.9891	< 2.2E-16
Lang$_{ij}$	0.8322413 **	18.5156	< 2.2E-16
Border$_{ij}$	1.1160283 **	11.4541	< 2.2E-16
Colony$_{ij}$	0.5963355 **	5.3953	6.852E-08
CU$_{ij}$	0.2120128 *	2.0001	0.0454926
Bid Host	0.0189928	1.4515	0.1466461
Bid Candidate	0.0164767	1.5318	0.1255809
Observations	122,747		
F-Statistics	237.05		0.00
R-Squared	0.8411		

Although, all the mega events' organizers (SOG, WOG, and FIFA WC) have certain qualification criteria for participation in the bidding process, which means if a country is deemed to be eligible for participation in the bidding, it is developed enough to be considered as a touristic

attraction. However, if the country is not hosting the event, it may not benefit from the economic gains. Restrictedness of the bid evaluation processes could be one of the reasons for this result. Reports on the bid evaluation for the three mega events included in the analysis outline a process where committees comprised of limited number of people hold meetings in order to finalize the winning bid. Another major factor in the tourism industry is publicity and media coverage, because of the significant marketing costs associated with it, only the countries who win the bid invest in it, and therefore benefit from the gains.

SECTION III: THE RISK INTEGRATED APPROACH TOWARDS IMPACT ANALYSIS OF MEGA EVENTS ON TOURISM

The above-mentioned gravity models for analysis of "*the impact of mega events on tourism"* have significant deterministic characteristics; however, one of the very important factors that shapes the visitor's state of mind in the decision making process from security, fear, and perceived satisfaction perspectives is not included in Rose and Spiegel (2011) and Fourie and Santano-Gallego's (2011) models. This factor is referred to as the Risk-factor, which could be unveiled from the diversity in the methodological and perspectival scholarly approaches towards the subject in the literature.

Scholars have been studying perceived risk from engineering and psychometric perspectives to quantify its impact on general tourism since the 1990s, yet it is hard to find an agreed upon method (Boo & Gu, 2010; Yang & Nair, 2014). Different scholars have had different approaches towards identification, modeling and analysis of risk in their studies of domestic and international tourisms. For example, Sequeira and Nunes (2008) used the dynamic panel data analysis to study the impact of political risk on international tourism. Mohammed and Sookram (2015), Baker et al. (2002), Chesney-Lind and Lind (1986), and George and Swart (2012) addressed risk in the context of crime in international tourism. Some others have studied the impact of local or regional terrorism, government instability, and relationship between governance and tourism (Lepp & Gibson, 2003; Ritcher & Waugh, 1986). Research results of these studies, which are vastly diverse, reveal important facts about the nature of tourists' sensitivity towards perceived risk; hence appealing for inclusion of a risk factor in the Gravity Model for the impact analysis of mega events in international tourism context.

Importance of the risk accountability intensifies even further when occurrence of unfortunate events such as the causalities of Israeli athletes in 1972 Munich Olympic Games viewed by nearly 800 million people are taken into account (Lepp & Gibson, 2003; Sonmez & Graefe, 1998). Publicity and media coverage is also one of the most important aspects of mega events that turns the host cities into brands. This same reason of reach to massive audience makes mega events highly acquired targets for terrorist groups. Hence, the study of the impact of mega events on tourism introduces a unique approach of accounting for risk in a comprehensive way through

unification of several risk indicators studied separately by other scholars and incorporating them into the Gravity Model for potentially more accurate results.

DATA COLLECTION AND METHODOLOGY

The impact of mega sports events on international tourism has been the center of focus in this thesis. So far, the Gravity Model, popular among researchers, has been re-evaluated over an extended period of time with the largest number of data points. Inclusion of the risk factor, defined as "probability of an unfortunate event happening" by Pinto and Garvey (2010) in the control variables for the three dimensional analysis of mega events and tourism is the first attempt towards a more comprehensive Gravity Model in the literature. All other control variables in the model such as trade, GDP PC, POP, PPP, distance, and the dummy variables remain the same as in Fourie and Santano-Gallego's (2011) work. Data for the risk factor have been collected from the World Bank's data base which is available online for public use. The World Bank has categorized these data into two groups. The first is titled as World Risk Indicators (WRI) and encompasses individual country based natural hazards, epidemics, adult mortality rate, homicide rate, and poverty headcount ratio for around 137 countries. The second is named World Governance Indicators (WGI) and is a quantified measurement of the qualitative indications of country-specific corruption, terrorism, political stability, rule of law, and government effectiveness and quality. Although the former seems as a better representative of risk, the latter is used in this analysis for two reasons:

1. The WRI data are available in the form of a lump, rather than annual.
2. The WRI might be a useful source for the local governmental institutions, their developmental plans, and policies concerning the local people, however its impact on international tourism and specifically on those who are traveling to attend mega events is nearly inconsiderate. On the other hand, the WGI data set represents the overall image of the country and gives a better estimate for the level of personal safety, property security, comfort, and satisfaction.

Using Equation 4 and the OLS Model discussed, validated, and used in the previous sections of this thesis with the help of R-Statistical package the analysis is continued. Robust standard error in the data set has been clustered and the country and year fixed effects by country pairs are included in the model. The new Gravity Model is:

$$\begin{aligned} Ln\ Tou_{ijt} = {} & \beta_0 + \beta_1 LnTrade_{ijt} + \beta_2 LnGDPpc_{it} + \beta_3 LnGDPpc_{jt} + \beta_4 LnPOP_{it} \\ & + \beta_5 LnPOP_{jt} + \beta_6 LnPPP_{ijt} + \beta_7 LnDist_{ij} + \beta_8 LnORK_{it} \\ & + \beta_9 LnORK_{jt} + \beta_{10} Lang_{ij} + \beta_{11} Border_{ij} + \beta_{12} Colony_{ij} \\ & + \beta_{13} CU_{ij} + \acute{\eta} E_{it} + \gamma_i + \delta_j + \lambda_t + u_{ijt} \ldots\ldots\ldots (4) \end{aligned} \tag{4}$$

Most of the variables in Equation 4 are introduced in Chapter II, the only new variables included in the model are the Over All Risk (ORK) factor and its disaggregated individual indicators, such that:

ORK_{it}: Overall risk values of *i* in *t*.

ORK_{jt}: Overall risk values of *j* in *t*.

$Corruption_{it}$: Control of corruption of *i* in *t*.

$Corruption_{jt}$: Control of corruption of *j* in *t*.

$Govt\ Effec_{it}$: Government effectiveness values of *i* in *t*.

$Govt\ Effec_{jt}$: Government effectiveness values of *j* in *t*.

$Rgltry\ Qlty_{it}$: Regulatory quality of *i* in *t*.

$Rgltry\ Qlty_{jt}$: Regulatory quality of *j* in *t*.

$Rul\ of\ Law_{it}$: Rule of law values of *i* in *t*.

$Rul\ of\ Law_{jt}$: Rule of law values of *j* in *t*.

$Politcl\ Stab_{it}$: Political stability and terrorism of *i* in *t*.

$Politcl\ Stab_{jt}$: Political stability and terrorism of *j* in *t*.

Disaggregated factors of risk are composite governance indicators in the WGI data set and provide cross country measures for over 200 economies, as discussed below. Providentially, all

the countries included in this study as destinations and origins have been covered with respect to availability of data.

a) **Political Stability and Terrorism:** according to the World Bank's WGI website (www.worldbank.org), the political stability and absence of violence and terrorism (*Politcl Stab*) indicator measures perceptions of the possibility of politically motivated violence including terrorism. Some of the very important variables used to measure this factor with their respective data sources are: security risk rating (IJT), political terrorism, international tensions and terrorist threats (EIU), armed conflicts (EIU), protests and riots (WMO), terrorism (WMO), interstate war (WMO), civil war (WMO), and intensity of violent activities (IPD).
b) **Control of Corruption:** The WGI defines this indicator as measure of the individual country based governmental corruption. This indicator (*Corruption*) captures different levels and forms of corruption by individuals in the public sector for private interests. Several imperative data sources are used for this measurement among which, the most important variables specifying overall measurement are listed by WGI as, corruption among public officials (EIU), irregular payments in export and import (GCS), whether corruption in government widespread (GWP), corruption (PRS) (WMO), irregular payments in judicial decisions (GCS), and level of "petty" corruption between administration and citizens (IPD).
c) **Rule of Law:** This indicator (*Rul of Law*) is defined as the measurement of perceptions to the extent to which citizens abide by the rules, quality of property rights, police, courts, and the likelihood of crime and violence happening. WGI's breakdown of the variables for this indictor include but are not limited to organized crime (EIU), violent crime (EIU), reliability of police services (GCS), confidence in the police forces (GWP), confidence in judicial system (GWP), degree of security of goods and persons by criminal organizations (drug trafficking, weapons, prostitution) (IPD), equal treatment of foreigners before the law (compared to nationals) (IPD), HER property rights (HER), law and order (PRS), and trafficking in people (TPR).
d) **Regulatory Quality:** The measurement of perceived governmental capabilities for formation and implementation of frame works and policies for the private sector's development and promotion are captured through this indicator (*Rgltry Qlty*). In addition to several other variables considered in specifying this indicator, some of the tourism influential factors could be named as price controls (EIU), excessive protections (EIU), prevalence of trade barriers

(GCS), effectiveness of anti-trust policy (GCS), investment freedom (HER), ease of setting up a subsidiary for a foreign firm (IPD), and regulatory burden (WMO).

e) **Government Effectiveness:** This indicator is described as "the perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (WGI, n.p.). The variables that are very relevant and fall under the tourism influential category of our study are: infrastructure (GCS), satisfaction with public transportation system (GWP), satisfaction with roads and highways (GWP), coverage area: basic health services (IPD), coverage area: electricity grid (IPD), coverage area: drinking water and sanitation (IPD), bureaucratic quality (PRS), state failure (WMO), and infrastructure disruption (WMO).

f) **Overall Risk Factor:** Although statistical analysis of the model in Section II allows for inclusion of around 200 additional variables, however, to keep the model simple and useful, an overall risk factor (*ORK*) is calculated by taking average of the five individual indicators. This process makes the model easy on the eyes, yet covers influence of all the factors supposedly having impact on the tourism.

The indicators values range from -2.5 to +2.5 with higher values indicating higher stability and betterment. In order to use the natural logarithm in the model used in this thesis, the negative sign was eliminated by scaling the range through adding 2.5 to the original values, changing the scale to 0.0001 (lowest) to 5.0000 (highest) ranking.

Empirical Results

Risk Integrated Impact of Mega Sports Events on Tourism

In this part of the study, the hypothesis of whether mega sports events increase tourism when the control variable Risk (ORK) is included in the model is tested. The analysis results of Equation 4 over the extended data set are presented in Table 10 below.

Table 10

Tourism and Mega Sports Events- Risk Integrated

	Estimate	T-Value	Pr(>|t|)
(Intercept)	-1.4356263	-0.7482	0.4543704
Ln $Trade_{ij}$	0.2855331 **	32.8500	< 2.2E-16
Ln $GDPPC_j$	0.1877017 **	6.9863	2.837E-12
Ln $GDPPC_i$	0.2316447 **	9.8641	< 2.2E-16
Ln POP_j	0.0248276	0.2913	0.7708320
Ln POP_i	0.4153795 **	4.6964	2.651E-06
Ln PPP_{ij}	-0.1019905 **	-7.8036	6.063E-15
Ln $Dist_{ij}$	-1.0837288 **	-44.2682	< 2.2E-16
$Lang_{ij}$	0.8324275 **	18.4402	< 2.2E-16
$Border_{ij}$	1.1186652 **	11.4165	< 2.2E-16
$Colony_{ij}$	0.5944263 **	5.3626	8.219E-08
CU_{ij}	0.2100925 *	1.9872	0.0469054
Ln ORK_j	0.0157812	0.1960	0.8446400
Ln ORK_i	1.0838589 **	13.7330	< 2.2E-16
Event	0.0377819 **	3.4478	0.0005653
Observations	121,888		
F-Statistics	236.52		0.00
R-Squared	0.8421		

Table 10 shows that the R-Squared value of the model has slightly increased to almost 85%. This is a sign of improvement in the overall performance of the model for better results. It is also apparent from Table 10 that coefficients' estimates and significance level of most of the control

variables remain consistent with the Section II results. The variable of main interest in this table is the Overall Risk Factor (ORK). Results in Table 10 show that the Overall Risk Factor is significantly positive for the tourism destination country, or more specifically the host country to the mega event. This means the higher the country's ORK ranking, the safer tourists feel when considering traveling there. The relatively large coefficient estimate of the variable calls for intrinsic attention to be paid to the inclusions of ORK when analyzing the impact of mega events on tourist arrivals or forecasting international tourism in general. On the other hand, the ORK of the tourism origin country is statistically insignificant. This result further confirms the validity of the extended model and the ORK estimates based on the practical logic that, regardless of the current situation of the origin country, tourists are concerned for their safety, security of their property, and facing unexpected situations at the destination country. Significance level of the population destination variable also increases from α=0.90 to α=0.99, implying direct relationship between population of destination and the ORK. Moreover, the statistical significance of the mega events also increases with its coefficient maintaining its original value.

Impact of Disaggregated Risk Indicators on Tourism

Although the overall risk factor is a holistic way of bringing all the different types of risk indicators in the literature into one place, the analysis of the five individual factors is also tempting. Most of the researchers in the literature have a conceived risk factor of their perception in conjunction with one or more of the controlled variables into their analysis of its impact on general tourism. The conclusions drawn in these empirical studies are largely based on smaller data samples. This is why the opportunity is utilized performing analysis of the disaggregated risk factors, so that the bias in smaller data samples is accounted for and suggestions are made with high level of confidence.

Table 11

Individual Risk Indicators' Impact on International Inbound Tourism

	Estimate	T-Value	Pr(>|t|)
(Intercept)	-1.92932	-1.0128	0.3111530
Ln Trade$_{ij}$	0.285808 **	32.9020	< 2.2E-16
Ln GDPPC$_j$	0.198546 **	7.5358	4.887E-14
Ln GDPPC$_i$	0.216055 **	9.1623	< 2.2E-16
Ln POP$_j$	0.028057	0.3311	0.7405346
Ln POP$_i$	0.44707 **	5.0927	3.535E-07
Ln PPP$_{ij}$	-0.08067 **	-6.4616	1.040E-10
Ln Dist$_{ij}$	-1.08344 **	-44.2831	< 2.2E-16
Lang$_{ij}$	0.831775 **	18.4262	< 2.2E-16
Border$_{ij}$	1.118128 **	11.4120	< 2.2E-16
Colony$_{ij}$	0.594848 **	5.3658	8.074E-08
CU$_{ij}$	0.209265 *	1.9792	0.0477945
Event	0.032659 **	2.9999	0.0027012
Ln Corruption$_j$	-0.03618	-0.5754	0.5650182
Ln Govt Effec$_j$	0.11779	1.5251	0.1272286
Ln Rgltry Qlty$_j$	-0.10705	-1.7112	0.0870400
Ln Politcl Stab$_j$	0.027838	1.3224	0.1860243
Ln Rul of Law$_j$	0.010716	0.1382	0.8900821
Ln Corruption$_i$	-0.17713 **	-2.9694	0.0029842
Ln Govt Effec$_i$	1.18382 **	16.5956	< 2.2E-16
Ln Politcl Stab$_i$	0.184506 **	6.9598	3.426E-12
Ln Rgltry Qlty$_i$	-0.13323 *	-2.3910	0.0168043
Ln Rul of Law$_i$	0.132903	1.8813	0.0599403
Observations	121,888		
F-Statistics	233.41		0.00
R-Squared	0.8421		

Table 11 reveals some very important facts about the individual factors. The results show that in addition to all individual factors of risk being statistically insignificant for the country of origin, the rule of law for the destination is also statistically not significant. Regulatory quality, political stability, government effectiveness, and corruption are all significant for the country of destination. An interesting result in Table 11, however, is the negative sign for corruption and regulatory quality variables. Although this is unexpected and needs further research for the cause-and-effect

analysis; still, when looked into the definitions of the disaggregated variables, the two positively significant factors have predominantly stronger accountability for dealing with foreigners than those with negative signs.

Participation in Bidding for Hosting Mega Event

Earlier in Section II the impact of participation in the bidding process was analyzed. With the inclusion of ORK in the model and introduction of Equation 4, the hypothesis of participation in bidding for a mega sport event and its relationship with increase in tourism is tested one more time.

Table 12

Risk Integrated Analysis of Participation in the Bidding Process

	Estimate	T-Value	Pr(>|t|)
(Intercept)	-1.4433607	1.9189655	0.4519589
Ln Trade$_{ij}$	0.2855414 **	0.0086920	< 2.2E-16
Ln GDPPC$_j$	0.1876380 **	0.0268693	2.897E-12
Ln GDPPC$_i$	0.2321288 **	0.0234921	< 2.2E-16
Ln POP$_j$	0.0247623	0.0852580	0.7714804
Ln POP$_i$	0.4157962 **	0.0884521	2.594E-06
Ln PPP$_{ij}$	-0.1020577 **	0.0130705	5.844E-15
Ln Dist$_{ij}$	-1.0837186 **	0.0244810	< 2.2E-16
Lang$_{ij}$	0.8324127 **	0.0451417	< 2.2E-16
Border$_{ij}$	1.1186759 **	0.0979870	< 2.2E-16
Colony$_{ij}$	0.5944263 **	0.1108472	8.219E-08
CU$_{ij}$	0.2101109 *	0.1057240	0.0468856
Ln ORK$_j$	0.0157702	0.0805297	0.8447428
Ln ORK$_i$	1.0825028 **	0.0789716	< 2.2E-16
Bid Host	0.0202292	0.0132440	0.1266584
Bid Candidate	0.0047155	0.0108185	0.6629326
Observations	121,888		
F-Statistics	236.03		0.00
R-Squared	0.8421		

Table 12 shows that the bid candidacy. or to be more precise, the entire bid process has no impact on the tourism until the event takes place in the destination country. However, it is strongly suggested for the organizers of mega events to include Risk factor in their qualification criteria of participation in the bidding process. Furthermore, if intended for the bidding process to be influential in tourist attraction, more publicity and inducing fundamental changes in the course of bidding may lead there.

Seasonality and Participation in Games and Its Impact on Tourism

Finally, the last hypothesis tested to complete the analysis in this thesis is whether the touristic impact of seasonality and participation versus no participation in the mega sports events changes with the new model. Results of the analysis are presented in Table 13.

Table 13

Risk Integrated Seasonality and Event participation Impact Analysis

	Estimate	T-Value	Pr(>|t|)	Estimate	T-Value	Pr(>|t|)
(Intercept)	-1.4596552	-0.7609	0.4466945	-1.4823057	-0.7725	0.4398085
Ln Trade$_{ij}$	0.2855104 **	32.8475	< 2.2E-16	0.2855627 **	32.8551	< 2.2E-16
Ln GDPPC$_j$	0.1882265 **	7.0090	2.414E-12	0.1877592 **	6.9902	2.759E-12
Ln GDPPC$_i$	0.2319736 **	9.8833	< 2.2E-16	0.2309541 **	9.8348	< 2.2E-16
Ln POP$_j$	0.0274005	0.3217	0.7477152	0.0268499	0.3152	0.7526178
Ln POP$_i$	0.4136838 **	4.6786	2.892E-06	0.4167110 **	4.7114	2.463E-06
Ln PPP$_{ij}$	-0.1020133 **	-7.8049	6.002E-15	-0.1024256 **	-7.8397	4.554E-15
Ln Dist$_{ij}$	-1.0837576 **	-44.2690	< 2.2E-16	-1.0836329 **	-44.2640	< 2.2E-16
Lang$_{ij}$	0.8323718 **	18.4389	< 2.2E-16	0.8321971 **	18.4371	< 2.2E-16
Border$_{ij}$	1.1187452 **	11.4175	< 2.2E-16	1.1185331 **	11.4158	< 2.2E-16
Colony$_{ij}$	0.5943819 **	5.3623	8.230E-08	0.5938059 **	5.3570	8.478E-08
CU$_{ij}$	0.2101854 *	1.9880	0.0468139	0.2100499 *	1.9868	0.0469449
ORK$_{jt}$	0.0162064	0.2014	0.8404180	0.0161555	0.2006	0.8409854
ORK$_{it}$	1.0837714 **	13.7341	< 2.2E-16	1.0822579 **	13.7138	< 2.2E-16
Peak Season	-0.0967258 **	-5.8406	5.214E-09			
Off Season	0.1108600 **	5.9267	3.099E-09			
Event part				0.1661551 **	7.2017	5.980E-13
Event no part				-0.0309018	-1.3445	0.1787832
Observations	121,888			121,888		
F-Statistics	236.01	0.00		236.24	0.00	
R-Squared	0.8421			0.8421		

With inclusion of ORK in the analysis, there is no considerable change in the behavior of the variables. This not only adds to the accuracy of results is Section II, it also is enabling to further elaborate on the cause and effect analysis of the seasonality factor. Crowd management, relationship between tourism and crime, infrastructure, transportation, exposure to terrorism, the host country's public sector's capabilities and services quality are the explanatory variables of ORK and important decision making criteria for travelers. Therefore, when a mega sports event is held in tourism peak season, the tourists' intentions are inevitably affected by the degree of seriousness of one or more of such factors leading to unexpectedly low outcomes. To solve this dilemma, the idealistic solution will be to strictly consider hosting mega events in off-peak seasons; in case this is not possible, then coordination of the location and provision of assurance to the targeted group of people on all the aforementioned variables is the alternative for better results. The analysis results for the differentiation in impact on tourism of participating and None-Participating countries in the mega events does not change either. Thus, due to the fact that host countries struggle to give the visitor a lasting impression for their visit, it is suggested to focus more on the tourists from the countries participating in the event.

CHAPTER IV

CONCLUSIONS AND FUTURE RESEARCH

The cost and benefit based empirical study of mega sport events in the international tourism context in this thesis grasps on the root causes for enhancement of expected economic gains and extends towards suggestion of a comprehensive model for analysis purposes. In the process of validation, Fourie and Santano-Gallego's (2011) work in is revisited. Changes are captured and important observations are made. The first two sections of this thesis focuses on validation of data extension, use of tools and methodology for analysis. The empirical results suggest that bilateral trade increases inbound tourist arrival. It also suggests that people from richer countries travel more often compared to nations that have lower income (GDP PC).

Moreover, difference in the prices (PPP) significantly influences the travel intentions of the people. The dummy variables included in the model account for the cultural and geographical aspects of the international inbound tourism where the results suggest that, people are more likely to visit counterpart countries sharing with them a common language, common border, common currency, colonial ties, or a combination of two or more of these variables. This means the cultural factor is highly important and therefore extreme measures should be taken to decide on the mega event's host country in a way that can incorporate as many of these aspects as possible, hence leading to an optimal outcome.

The results also emphasize on independent studies of the interested countries, in case they find themselves meeting most of the criterion suggested in this thesis, they are encouraged to present stronger bids so that they can host the mega events and benefit from the economic gains. The findings in this study show that population of the destination country could be a plus point in the attraction of international tourism, especially with the current era of global connectivity and social media generation. The focus on mega events in this thesis shows that the tourism gains will significantly depend on the popularity-density of the event, less popular events such as the Rugby Lions tours do not have much impact on international tourism enhancement.

The historical data show that tourism gains could be achieved one year prior to the event and in the year of event. However, continued publicity and achievement of visitor's satisfaction during the visit may change this trend. The time of the mega event is found to be very important and it is

suggested for the organizing committees to plan the events for an off-peak tourism season. In this way, not only will the regular general purpose tourists continue to visit, but the host country will also receive an increased number of new tourists, specifically for the mega event. It is also suggested that the organizing committees coordinate two or more mega events happening in the same year. In the ideal case, it is recommended to host the mega events in different years, however if this is not possible, then a concentration on the geographical distance and overlapping nations can be a good way of accounting for this factor. The research findings in this thesis show that participation in the bidding process for hosting such mega events should not be interpreted as resulting in significant tourism gains. Moreover, the host countries should give special consideration in their preparation to inbound tourists from countries participating in the events.

Finally, the risk control variable is introduced and the Gravity Model is extended for increased accuracy and further comprehensiveness of the analytics. Results of the analysis with the extended model shows significant importance of the variable, and validates the argument of risk being a substantial factor in shaping the tourist's final decision. The analysis of the disaggregated risk factor provides clarification to the controversies over the important and not important aspects of risk in the literature. Analysis results in Section III show that government effectiveness and political stability (each accounting for further disaggregated factors) remain on the top of the list, and rule of law is statistically insignificant. Although the negative signs of the control of corruption and regulatory quality variables are justified based on the available information, researchers are encouraged to further explore the cause and effect.

FUTURE RESEARCH

Although, the research results in this thesis are based on evidently one of the largest number of observations and most up to date data set in the literature, the areas listed below are left for future research and further exploration by interested researchers.

1. Performing the best-subset regression as a continuation to the current study.
2. Performing an in-depth study of the event's lasting legacy effects for three or more years before and after the event and potential causes for the negative signs.
3. Performing an in-depth study of the disaggregated risk indicators investigating for the reasons behind the negative signs for control variables of corruption and regulator quality variables.

4. Inclusion of a global risk factor in the model to account for impact of global risk on the tourists' mindset and perceived risk.

REFERENCES

Akerloff, G., Rose, A., Yellen, J., & Hessenius, H. (1991). East Germany in from the Cold: The Economic Aftermath of the Currency Union. *Brookings Papers on Economic Activity, 1(1)*, 1-101.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting, 27*(3), 870-886.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, *93*(1), 170-192.

Barker, M., Page, S. J., & Meyer, D. (2002). Modeling tourism crime: The 2000 America's cup. *Annals of Tourism Research*, *29*(3), 762-782.

Birchenall, J. A. (2016). Population and Development Redux. *Journal of Population Economics*, *29*(2), 627-656. doi:http://dx.doi.org/10.1007/s00148-015-0572-x

Boo, S., & Gu, H. (2010). Risk perception of mega-events. *Journal of Sport & Tourism*, *15*(2), 139-161.

Chesney-Lind, M., & Lind, I. Y. (1986). Visitors as victims: Crimes against tourists in Hawaii. *Annals of Tourism Research*, *13*(2), 167-191.

Claveria, O., Monte, E., & Torra, S. (2015). Tourism demand forecasting with neural network models: different ways of treating information. *International Journal of Tourism Research*, *17*(5), 492-500.

Costantini, M., & Pappalardo, C. (2010). A hierarchical procedure for the combination of forecasts. *International Journal of Forecasting*, *26*(4), 725-743.

Crouch, G. I. (1994). The study of international tourism demand: A survey of practice. *Journal of Travel Research*, *32*(4), 41-55.

De Sousa, J., Mirza, D., & Verdier, T. (2010). Terrorism networks and trade: Does the neighbor hurt?. *IDEAS Working Paper Series from RePEc.*

Dornbusch, R. (1991). East Germany in from the cold: The economic aftermath of currency union: Comments. *Brookings Papers on Economic Activity*, *1*, 88-92.

Eilat, Y., & Einav, L. (2004). Determinants of international tourism: A three-dimensional panel data analysis. *Applied Economics*, *36*(12), 1315-1327.

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

George, R., & Swart, K. (2012). International tourists' perceptions of crime–risk and their future travel intentions during the 2010 FIFA World Cup™ in South Africa. *Journal of Sport & Tourism*, *17*(3), 201-223.

Gursoy, D., & Kendall, K. W. (2006). Hosting mega events: Modeling locals' support. *Annals of Tourism Research*, *33*(3), 603-623.

Hanafiah, M. H. M., & Harun, M. F. M. (2010). Tourism Demand in Malaysia: A cross-sectional pool time-series analysis *International Journal of Trade, Economics and Finance*, *1*(1), 80.

Hayes, G., & Karamichas, J. (2011). *Olympic Games, mega-events, and civil societies: Globalisation, environment, resistance*. City: Palgrave Macmillan.

Hiller, H. H. (1995). Conventions as mega-events: A new model for convention-host city relationships. *Tourism Management*, *16*(5), 375-379.

Hiller, H. H. (1998). Assessing the impact of mega-events: A linkage model. *Current Issues in Tourism*, *1*(1), 47-57.

Horne, J., & Manzenreiter, W. (2006). An introduction to the sociology of sports mega-events1. *The Sociological Review*, *54*(s2), 1-24.

International Monetary Fund. (n.d.). Direction of trade statistics. Retrieved from IMF website: http://data.imf.org/regular.aspx?key=61013712

Jensen, C., & Zhang, J. (2013). Trade in tourism services: Explaining tourism trade and the impact of the general agreement on trade in services on the gains from trade. *The Journal of International Trade & Economic Development*, *22*(3), 398-429.

Keum, K. (2010). Tourism flows and trade theory: A panel data analysis with the gravity model. *The Annals of Regional Science*, *44*(3), 541-557.

Kim, J., & Lee, C. K. (2016). Role of tourism price in attracting international tourists: The case of Japanese inbound tourism from South Korea. *Journal of Destination Marketing & Management (Article in Press)*. http://dx.doi.org/10.1016/j.jdmm.2016.03.002i

Kim, N. S., & Chalip, L. (2004). Why travel to the FIFA World Cup? Effects of motives, background, interest, and constraints. *Tourism Management*, *25*(6), 695-707.

Lee, C. K., Taylor, T., Lee, Y. K., & Lee, B. (2005). The impact of a sport mega-event on destination image: The case of the 2002 FIFA World Cup Korea/Japan. *International Journal of Hospitality & Tourism Administration*, *6*(3), 27-45.

Lepp, A., & Gibson, H. (2003). Tourist roles, perceived risk and international tourism. *Annals of Tourism Research*, *30*(3), 606-624.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82-99.

Lim, C. (1999). A meta-analytic review of international tourism demand. *Journal of Travel Research*, *37*(3), 273-284.

Maennig, W., & Zimbalist, A. S. (Eds.). (2012). *International handbook on the economics of mega sporting events*. Edward Elgar Publishing.Mayer, T., & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database.

Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, *93*(2), 351-363.

Mohammed, A., & Sookram, S. (2015). The impact of crime on tourist arrivals--A comparative analysis of Jamaica and Trinidad and Tobago. *64*(2), 153-176

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research*, *48*, 1-10.

Müller, M. (2015). What makes an event a mega-event? Definitions and sizes. *Leisure Studies*, *34*(6), 627-642.

Narayan, S., & Nguyen, T. T. (2016). Does the trade gravity model depend on trading partners? Some evidence from Vietnam and her 54 trading partners. *International Review of Economics & Finance*, *41*, 220-237.

Nyikana, S., Tichaawa, T. M., & Swart, K. (2014). Sport, tourism and mega-event impacts on host cities: A case study of the 2010 FIFA World Cup in Port Elizabeth.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, 181-193.

Pinto, C. A., & Garvey, P. R. (2012). *Advanced risk analysis in engineering enterprise systems*. City: CRC Press.

Ritchie, J. B., & Yangzhou, J. (1987, August). The role and impact of mega-events and attractions on national and regional tourism: a conceptual and methodological overview. In *Proceedings of the 37th Congress of AIEST* (Vol. 28, pp. 17-57).

Richter, L. K., & Waugh, W. L. (1986). Terrorism and tourism as logical companions. *Tourism Management*, *7*(4), 230-238.

Roche, M. (2000). *Megaevents and modernity: Olympics and expos in the growth of global culture*. City: Routledge.

Rose, A. K. (2000). One money, one market: the effect of common currencies on trade. *Economic Policy*, *15*(30), 08-45.

Rose, A. K., & Spiegel, M. M. (2011). The Olympic effect. *The Economic Journal*, *121*(553), 652-677.

Santana-Gallego, M., Ledesma-Rodríguez, F. J., & Pérez-Rodríguez, J. V. (2016). International trade and tourism flows: An extension of the gravity model. *Economic Modelling*, *52*, 1026-1033.

Sequeira, T. N., & Nunes, P. (2008). Does country risk influence international tourism? A dynamic panel data analysis. *Economic Record*, *84*(265), 223-236.

Shi, Y., Yu, N. and Chen, R.J.C. (2015). After 2008 Olympic Games: From a Glorious Mega-event to a Sustainable Destination Brand. *Journal of Hospitality and Tourism, 15(1)*, 47-56.

Singha, K., & Jaman, M. S. (2013). Does population growth affecteconomic development? A study of India. *Journal of International Economics*, *4*(2), 41.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, *29*(2), 203-220.

Song, H., & Witt, S. F. (2000). *Tourism demand modelling and forecasting: Modern econometric approaches*. City: Routledge.

Sönmez, S. F., & Graefe, A. R. (1998). Influence of terrorism risk on foreign tourism decisions. *Annals of Tourism Research*, *25*(1), 112-144.

Surugiu, C., Leitão, N. C., & Surugiu, M. R. (2011). A panel data modelling of international tourism demand: Evidences for Romania. *Economic Research-Ekonomska Istraživanja*, *24*(1), 134-145.

The World Bank. (n.d.). World development indicators (wdi 2015). Retrieved from World Bank website: http://archive.data.worldbank.org/data-catalog/world-development-indicators/wdi-2015

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International journal of Forecasting*, *11*(3), 447-475.

World Bank Group (Ed.). (2015). *World Development Indicators 2015*. World Bank Publications.

World Tourism Organization Network (n.d.) Methodological notes to the tourism statistics database. Retrieved from: http://statistics.unwto.org/news/2015-03-05/methodological-notes-tourism-statistics-database

Yang, C. L., & Nair, V. (2014). Risk perception study in tourism: Are we really measuring perceived risk?. *Procedia-Social and Behavioral Sciences*, *144*, 322-327.

Youhan, S., Ning, Y., & Chen, R. C. (2015). After 2008 Olympic Games: From a glorious mega event to a sustainable destination brand. *Journal of Hospitality & Tourism*, *13*(1), 48-57.

APPENDICES

APPENDIX A. LIST OF COUNTRIES

Appendix A1

Table A1

List of Tourism Destination Countries

Albania	Congo	Iran, Islamic Republic of	Netherlands
Algeria	Congo (Democratic Republic of the)	Iraq	New Caledonia
Argentina	Cook Islands	Ireland	New Zealand
Armenia	Costa Rica	Israel	Nicaragua
Aruba	Cote d'Ivoire	Italy	Niger
Australia	Croatia	Jamaica	Nigeria
Austria	Cuba	Japan	Norway
Azerbaijan	Cyprus	Kazakhstan	Oman
Bahamas	Czech Republic	Kenya	Panama
Bahrain	Denmark	Korea, Republic of	Papua New Guinea
Bangladesh	Dominica	Kuwait	Paraguay
Barbados	Dominican Republic	Kyrgyz Republic	Peru
Belarus	El Salvador	Lao People's Democratic Republic	Poland
Belgium	Estonia	Latvia	Portugal
Belize	Ethiopia	Libya	Puerto Rico
Benin	Fiji	Liechtenstein	Romania
Bermuda	Finland	Lithuania	Russian Federation
Bolivia	France	Luxembourg	Rwanda
Bosnia and Herzegovina	Gabon	Macedonia, FYR	Saint Kitts and Nevis
Brazil	Gambia	Madagascar	Saint Lucia
British Virgin Islands	Georgia	Malawi	Saint Vincent and the Grenadines
Brunei Darussalam	Germany	Malaysia	Sao Tome and Principe
Bulgaria	Greece	Maldives	Senegal
Burkina Faso	Grenada	Mali	Serbia
Cambodia	Guatemala	Malta	Seychelles
Cameroon	Guinea	Mauritius	Singapore
Canada	Guinea-Bissau	Mexico	Slovak Republic
Cape Verde	Haiti	Moldova	Slovenia

Table A1. (continued)

Central African Republic	Honduras	Monaco	Solomon Islands
Chad	Hong Kong	Mongolia	South Africa
Chile	Hungary	Montserrat	Spain
China	Iceland	Morocco	Sri Lanka
Colombia	India	Mozambique	Sudan
Comoros	Indonesia	Nepal	Suriname
Sweden	Togo	Turks and Caicos	Vanuatu
Switzerland	Tonga	Uganda	Venezuela
Syrian Arab Republic	Trinidad and Tobago	United Arab Emirates	Vietnam
Tajikistan	Tunisia	United Kingdom	Yemen
Tanzania	Turkey	United States	Zambia
Thailand	Turkmenistan	Uruguay	Zimbabwe

Appendix A2

Table A2 List of Tourism Origin Countries

Afghanistan	Chad	Ghana	Lesotho
Albania	Chile	Gibraltar	Liberia
Algeria	China	Greece	Libya
Angola	Colombia	Greenland	Lithuania
Antigua and Barbuda	Comoros	Grenada	Luxembourg
Argentina	Congo	Guatemala	Macao
Armenia	Congo (Democratic Republic of the)	Guinea	Macedonia, FYR
Aruba	Costa Rica	Guinea-Bissau	Madagascar
Australia	Cote d'Ivoire	Guyana	Malawi
Austria	Croatia	Haiti	Malaysia
Azerbaijan	Cuba	Honduras	Maldives
Bahamas	Cyprus	Hong Kong	Mali
Bahrain	Czech Republic	Hungary	Malta
Bangladesh	Denmark	Iceland	Mauritania
Barbados	Djibouti	India	Mauritius
Belarus	Dominica	Indonesia	Mexico
Belgium	Dominican Republic	Iran, Islamic Republic of	Moldova
Belize	Ecuador	Iraq	Mongolia
Benin	Egypt	Ireland	Morocco
Bermuda	El Salvador	Israel	Mozambique
Bhutan	Equatorial Guinea	Italy	Namibia
Bolivia	Eritrea	Jamaica	Nauru
Bosnia and Herzegovina	Estonia	Japan	Nepal
Botswana	Ethiopia	Jordan	Netherlands
Brazil	Falkland Islands	Kazakhstan	Netherlands Antilles
Brunei Darussalam	Faroe Islands	Kenya	New Caledonia
Bulgaria	Fiji	Kiribati	New Zealand
Burkina Faso	Finland	Korea, dem	Nicaragua
Burundi	France	Korea, Republic of	Niger
Cambodia	French Polynesia	Kuwait	Nigeria
Cameroon	Gabon	Kyrgyz Republic	Norway
Canada	Gambia	Lao People's Democratic Republic	Oman
Cape Verde	Georgia	Latvia	Pakistan
Central African Republic	Germany	Lebanon	Palau

Table A2. (continued)

Panama	Saint Vincent and the Grenadines	Sri Lanka	Turkmenistan
Papua New Guinea	Samoa	Sudan	Uganda
Paraguay	Sao Tome and Principe	Suriname	Ukraine
Peru	Saudi Arabia	Swaziland	United Arab Emirates
Philippines	Senegal	Sweden	United Kingdom
Poland	Serbia	Switzerland	United States
Portugal	Seychelles	Syrian Arab Republic	Uruguay
Qatar	Sierra Leone	Tajikistan	Uzbekistan
Romania	Singapore	Tanzania	Vanuatu
Russian Federation	Slovak Republic	Thailand	Venezuela
Rwanda	Slovenia	Togo	Vietnam
Saint Helena	Solomon Islands	Tonga	Yemen
Saint Kitts and Nevis	Somalia	Trinidad and Tobago	Zambia
Saint Lucia	South Africa	Tunisia	Zimbabwe
Saint Pierre and Miquelon	Spain	Turkey	

APPENDIX B. PARTICIPATION IN BIDDING PROCESS

Table B1

Participation in the Bidding Process

Year	Summer Olympic Games (SOG)	Winter Olympic Games (WOG)	FIFA World Cup (FIFA)
1996	USA(won) United Kingdom (lost) Australia (lost) Canada (lost)		
1998		Japan(won) Italy (lost) Spain (lost) Sweden (lost) USA (lost)	France(won) Morocco (lost) Switzerland (lost)
2000	Australia(won) China (lost) Germany (lost) Turkey (lost) United Kingdom (lost)		
2002		USA(won) Sweden (lost) Canada (lost) Switzerland (lost)	South Korea/Japan (won) Mexico (lost)
2004	Greece(won) Argentina (lost) South Africa (lost) Italy (lost) Sweden (lost)		
2006		Italy(won) Switzerland (lost) Finland (lost) Austria (lost) Slovak Rep. (lost)	Germany(won) South Africa (lost) Morocco (lost) Germany (lost) Brazil (lost)
2008	Canada (lost) China (won) France (lost) Japan (lost) Turkey (lost)	Poland(lost)	
2010		Austria (lost) Canada (won) South Korea (lost)	Egypt (lost) Morocco (lost) South Africa (won)

Table B1. (continued)

2012	France (lost)
	Russian Federation (lost)
	Spain (lost)
	United Kingdom (won)
	USA (lost)

APPENDIX C. THE R CODE

The R code used to perform OLS regression with robust Standard error being clustered. The results displayed in this example are that of Table 1.

```
> install.packages("sandwich")
> library(sandwich)
> install.packages("plm")
> library(plm)
> install.packages("multiwayvcov")
> library(multiwayvcov)
> library(lmtest)
> dsmse <- read.table("F:\\ DatasetMegaEvents.txt", header = TRUE, fill = TRUE)
> attach(dsmse)
> dsmse <- dsmse[complete.cases(dsmse),]
> names(dsmse)
      [1] "year"              "iddest"            "idorig"            "idpair"            "idyear"
      "lntou"           "lntrade"
       [8]  "lngdppcorig"          "lngdppcdest"          "lnpoporig"          "lnpopdest"
      "lngdporig"       "lngdpdest"       "Lnppporig"
      [15] "Lnpppdest"          "lnppp"              "lndist"             "CorruptionOrig"
      "GovtEffecOrig"     "PolitStaEstOrig"    "RegulatoryQualityOrig"
      [22] "RuleofLawOrig"     "OverallRiskOrig"    "CorruptionDest"     "GovtEffecDest"
      "PolitStabEstDest"   "RegulatoryQualityDest" "RuleofLawDest"
      [29] "OverallRiskDest"   "contig"          "comlang"         "colony"          "fta"
      "cu"              "event"
      [36] "SOG"             "WOG"             "FIFA"            "CWC"             "RWC"
      "lion"            "eventpart"
      [43] "eventnopart"     "event_l1"        "event_l2"        "event_l3"        "event_f1"
      "event_f2"        "event_f3"
      [50] "eventpeak"       "eventoffpeak"    "bidhost"         "bidcandidate"
> length(dsmse$year)
      [1] 83520
> ml <-

      lm(lntou~lntrade+lngdppcorig+lngdppcdest+lnpoporig+lnpopdest+lnppp+lndist+comlan
      g+contig+colony+cu+event+factor(iddest)+factor(idorig)+factor(idyear), data=dsmse)
> summary(ml)

      Call:
      lm(formula = lntou ~ lntrade + lngdppcorig + lngdppcdest + lnpoporig +
         lnpopdest + lnppp + lndist + comlang + contig + colony +
         cu + event + factor(iddest) + factor(idorig) + factor(idyear),
         data = dsmse)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-8.6520 -0.7577 -0.0116  0.7632  9.2131

Coefficients:
                Estimate  Std. Error   t value      Pr(>|t|)
(Intercept)    11.828297   2.341739     5.051    4.40e-07 ***
lntrade         0.069102   0.001276    54.160    < 2e-16 ***
lngdppcorig     0.268669   0.028175     9.536    < 2e-16 ***
lngdppcdest     0.170371   0.021317     7.992    1.34e-15 ***
lnpoporig       0.022858   0.133623     0.171    0.864172
lnpopdest      -0.074619   0.019268    -3.873    0.000108 ***
lnppp          -0.032236   0.008094    -3.983    6.82e-05 ***
lndist         -1.482243   0.007413  -199.944    < 2e-16 ***
comlang         1.075399   0.015866    67.780    < 2e-16 ***
contig          1.183954   0.026117    45.332    < 2e-16 ***
colony          0.918178   0.032640    28.130    < 2e-16 ***
cu              0.228876   0.043125     5.307    1.12e-07 ***
event           0.077874   0.035839     2.173    0.029791 *

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.308 on 83172 degrees of freedom
Multiple R-squared:  0.8376,    Adjusted R-squared:  0.8369
F-statistic:  1236 on 347 and 83172 DF,  p-value: < 2.2e-16

> get_CL_vcov <- function(model, cluster){
+ require(sandwich, quietly=TRUE)
+ require(lmtest, quietly=TRUE)
+ M <- length(unique(cluster))
+ N <- length(cluster)
+ K <- model$rank
+ dfc <- (M/(M - 1))*((N - 1)/(N - K))
+ uj <- apply(estfun(model), 2, function(x) tapply(x, cluster, sum))
+ vcovCL <- dfc*sandwich(model, meat=crossprod(uj)/N)
+ return(vcovCL)
+ }
> ml.vcovCL <- get_CL_vcov(ml, dsmse$idpair)

> coeftest(ml, ml.vcovCL)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.8282975	3.0220601	3.9140	9.086e-05 ***
lntrade	0.0691017	0.0035135	19.6676	< 2.2e-16 ***
lngdppcorig	0.2686685	0.0327646	8.2000	2.438e-16 ***
lngdppcdest	0.1703710	0.0321083	5.3061	1.123e-07 ***
lnpoporig	0.0228584	0.1721572	0.1328	0.8943706
lnpopdest	-0.0746186	0.0339218	-2.1997	0.0278291 *
lnppp	-0.0322355	0.0121477	-2.6536	0.0079646 **
lndist	-1.4822426	0.0231776	-63.9514	< 2.2e-16 ***
comlang	1.0753988	0.0494830	21.7327	< 2.2e-16 ***
contig	1.1839538	0.1111004	10.6566	< 2.2e-16 ***
colony	0.9181783	0.1144333	8.0237	1.039e-15 ***
cu	0.2288761	0.1209245	1.8927	0.0583988 .
event	0.0778742	0.0123936	6.2834	3.328e-10 ***

```
> get_confint <- function(model, vcovCL){
+ t <- qt(0.975, model$df.residual)
+ ct <- coeftest(model, vcovCL)
+ est <- cbind(ct[,1], ct[,1]-t*ct[,2], ct[,1]+t*ct[,2])
+ colnames(est) <- c("Estimate", "LowerCI", "UpperCI")
+ return(est)
+ }
> get_confint(ml, ml.vcovCL)
```

	Estimate	LowerCI	UpperCI
(Intercept)	11.828297473	5.905082309	17.751512636
lntrade	0.069101708	0.062215308	0.075988109
lngdppcorig	0.268668504	0.204450152	0.332886856
lngdppcdest	0.170371034	0.107439051	0.233303018
lnpoporig	0.022858398	-0.314568360	0.360285155
lnpopdest	-0.074618623	-0.141105082	-0.008132163
lnppp	-0.032235526	-0.056044927	-0.008426125
lndist	-1.482242615	-1.527670622	-1.436814609
comlang	1.075398778	0.978412438	1.172385119
contig	1.183953788	0.966197831	1.401709745
colony	0.918178282	0.693889930	1.142466633
cu	0.228876064	-0.008135071	0.465887199
event	0.077874156	0.053582887	0.102165425

```
> waldtest(ml, vcov = ml.vcovCL, test = "F")
```

Wald test

Model 1: lntou ~ lntrade + lngdppcorig + lngdppcdest + lnpoporig + lnpopdest + lnppp + lndist + comlang + contig + colony + cu + event +

```
    factor(iddest) + factor(idorig) + factor(idyear)
Model 2: lntou ~ 1
  Res.Df     Df           F        Pr(>F)
1  83172
2  83519  -347  213.63   < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

APPENDIX D. CONFIDENCE INTERVALS

Table D1

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	11.828297473	5.905082309	17.751512636	11.73853863	5.816814049	17.66026321
Ln Trade$_{ij}$	0.069101708	0.062215308	0.075988109	0.069098179	0.062211606	0.075984753
Ln GDPPC$_{j}$	0.268668504	0.204450152	0.332886856	0.268570026	0.204371505	0.332768547
Ln GDPPC$_{i}$	0.170371034	0.107439051	0.233303018	0.171123089	0.108120790	0.234125389
Ln POP$_{j}$	0.022858398	-0.314568360	0.360285155	0.027470570	-0.309896571	0.364837712
Ln POP$_{i}$	-0.074618623	-0.141105082	-0.008132163	-0.074145286	-0.140562128	-0.007728444
Ln PPP$_{ij}$	-0.032235526	-0.056044927	-0.008426125	-0.032606264	-0.056421860	-0.008790668
Ln Dist$_{ij}$	-1.482242615	-1.527670622	-1.436814609	-1.482256713	-1.527686189	-1.436827238
Lang$_{ij}$	1.075398778	0.978412438	1.172385119	1.075425687	0.978436439	1.172414935
Border$_{ij}$	1.183953788	0.966197831	1.401709745	1.183861853	0.966095916	1.401627790
Colony$_{ij}$	0.918178282	0.693889930	1.142466633	0.918188663	0.693893165	1.142484161
CU$_{ij}$	0.228876064	-0.008135071	0.465887199	0.228606380	-0.008424821	0.465637580
Event	0.077874156	0.053582887	0.102165425			
SOG				0.177645556	0.111085828	0.244205284
WOG				-0.068440958	-0.108603666	-0.028278250
FIFA				0.076469231	0.022624164	0.130314297
CWC				0.192806756	0.126602588	0.259010924
RWC				-0.123215392	-0.198954210	-0.047476573
Lion				0.146157549	0.081392020	0.210923077

Table D2

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	11.773545385	5.849334379	17.6977563911	11.896450781	5.971149665	17.821751897
Ln Trade$_{ij}$	0.069112885	0.062225908	0.0759998616	0.069107000	0.062221011	0.075992989
Ln GDPPC$_j$	0.268809084	0.204590539	0.3330276288	0.268486270	0.204266527	0.332706014
Ln GDPPC$_i$	0.170303378	0.107342587	0.2332641693	0.170686970	0.107762439	0.233611501
Ln POP$_j$	0.026296529	-0.311171808	0.3637648657	0.018572693	-0.318972418	0.356117804
Ln POP$_i$	-0.074900584	-0.141492163	-0.0083090050	-0.074395826	-0.140804640	-0.007987012
Ln PPP$_{ij}$	-0.032429812	-0.056245572	-0.0086140516	-0.031894350	-0.055710295	-0.008078405
Ln Dist$_{ij}$	-1.482247539	-1.527676829	-1.4368182482	-1.482194810	-1.527624295	-1.436765325
Lang$_{ij}$	1.075216231	0.978229828	1.1722026353	1.075401263	0.978413763	1.172388763
Border$_{ij}$	1.183830834	0.966065115	1.4015965518	1.183954657	0.966200340	1.401708973
Colony$_{ij}$	0.917855582	0.693535097	1.1421760668	0.918201262	0.693912524	1.142490000
CU$_{ij}$	0.228518243	-0.008578104	0.4656145899	0.230061588	-0.006884100	0.467007277
Participant	0.117594526	0.059708277	0.1754807742			
Non Participant	0.003056706	-0.051653103	0.0577665160			
Peak Season				-0.030232355	-0.081354715	0.020890004
Off Season				0.167231346	0.123685197	0.210777496

Table D3

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	6.77125123	-0.0704278958	13.61293036	6.558030078	-2.856324E-01	13.401692542
Ln $Trade_{ij}$	0.30149327	0.2832532805	0.31973325	0.301536731	2.832963E-01	0.319777182
Ln $GDPPC_j$	0.10726896	0.0389848390	0.17555308	0.107113684	3.886644E-02	0.175360932
Ln $GDPPC_i$	0.30535372	0.2436852474	0.36702219	0.306417985	2.446776E-01	0.368158357
Ln POP_j	-0.28893951	-0.5917356021	0.01385658	-0.282530659	-5.851817E-01	0.020120348
Ln POP_i	0.21859719	-0.0907602169	0.52795460	0.225217663	-8.426765E-02	0.534702973
Ln PPP_{ij}	-0.05229313	-0.0847894636	-0.01979680	-0.053285856	-8.578915E-02	-0.020782564
Ln $Dist_{ij}$	-1.08250522	-1.1329656758	-1.03204476	-1.082440801	-1.132902E+00	-1.031979634
$Lang_{ij}$	0.81641630	0.7228440026	0.90998860	0.816439352	7.228640E-01	0.910014669
$Border_{ij}$	1.10274551	0.9021432065	1.30334781	1.102651919	9.020395E-01	1.303264369
$Colony_{ij}$	0.67868570	0.4565998920	0.90077151	0.678765207	4.566920E-01	0.900838387
CU_{ij}	0.19250432	-0.0162724429	0.40128108	0.192353553	-1.642626E-02	0.401133366
Event	0.06800869	0.0431677340	0.09284964			
SOG				0.200941649	1.311953E-01	0.270688007
WOG				-0.077404392	-1.232177E-01	-0.031591113
FIFA				0.118700537	6.709758E-02	0.170303498
CWC				0.223235986	1.589842E-01	0.287487790
RWC				-0.142425003	-2.183656E-01	-0.066484413
Lion				0.033540715	-3.475887E-02	0.101840298

Table D4

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	6.88167085	0.0356932286	13.72764847	6.66645083	-1.774932E-01	13.510394853
Ln $Trade_{ij}$	0.30147544	0.2832355216	0.31971535	0.30153662	2.832960E-01	0.319777279
Ln $GDPPC_j$	0.10721300	0.0389250497	0.17550094	0.10737345	3.910058E-02	0.175646313
Ln $GDPPC_i$	0.30516814	0.2434763629	0.36685993	0.30512517	2.434480E-01	0.366802366
Ln POP_j	-0.29148309	-0.5944386565	0.01147248	-0.28489631	-5.876835E-01	0.017890927
Ln POP_i	0.21415353	-0.0953037895	0.52361084	0.22116210	-8.820365E-02	0.530527842
Ln PPP_{ij}	-0.05208334	-0.0846044395	-0.01956224	-0.05270533	-8.519521E-02	-0.020215440
Ln $Dist_{ij}$	-1.08253284	-1.1329941704	-1.03207152	-1.08247277	-1.132934E+00	-1.032011727
$Lang_{ij}$	0.81642445	0.7228520236	0.90999688	0.81628149	7.227081E-01	0.909854840
$Border_{ij}$	1.10278775	0.9021838554	1.30339165	1.10258894	9.019824E-01	1.303195482
$Colony_{ij}$	0.67868403	0.4565895510	0.90077851	0.67830382	4.561980E-01	0.900409671
CU_{ij}	0.19256169	-0.0162138303	0.40133720	0.19241326	-1.637840E-02	0.401204924
Peak Season	0.01073272	-0.0396662158	0.06113165			
Off Season	0.10188871	0.0554452523	0.14833217			
Participant				0.14258234	8.346972E-02	0.201694954
Non Participant				0.06980920	7.335421E-03	0.132282981

Table D5

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	2.745212834	-0.913148192	6.4035738597	2.689940E+00	-0.969392478	6.3492718921
Ln $Trade_{ij}$	0.285513962	0.268530987	0.3024969361	2.855182E-01	0.268535381	0.3025010636
Ln $GDPPC_j$	0.184942948	0.132509808	0.2373760867	1.848211E-01	0.132415675	0.2372265981
Ln $GDPPC_i$	0.302831439	0.256794631	0.3488682465	3.017489E-01	0.255674264	0.3478236219
Ln POP_j	-0.005420618	-0.173089192	0.1622479549	-4.078748E-03	-0.171658024	0.1635005292
Ln POP_i	0.191632752	0.023159814	0.3601056900	1.944635E-01	0.025948057	0.3629790145
Ln PPP_{ij}	-0.107216146	-0.134436334	-0.0799959571	-1.078551E-01	-0.135063716	-0.0806464619
Ln $Dist_{ij}$	-1.083870508	-1.132163358	-1.0355776568	-1.083855E+00	-1.132148836	-1.0355612868
$Lang_{ij}$	0.832257751	0.744159460	0.9203560432	8.322449E-01	0.744145019	0.9203447825
$Border_{ij}$	1.116001046	0.925031617	1.3069704744	1.115961E+00	0.924986583	1.3069356134
$Colony_{ij}$	0.596325748	0.379692390	0.8129591064	5.963814E-01	0.379753011	0.8130098509
CU_{ij}	0.211987803	0.004227426	0.4197481796	2.119064E-01	0.004149618	0.4196632448
Event	0.034030736	0.013053119	0.0550083540			
SOG				1.996257E-01	0.146896197	0.2523552630
WOG				-1.058111E-01	-0.142258075	-0.0693641559
FIFA				8.017667E-02	0.012195547	0.1481577921
CWC				1.680710E-01	0.117630067	0.2185119516
RWC				-1.192771E-01	-0.179239757	-0.0593145276
Lion				-3.078298E-02	-0.072158855	0.0105929018

Table D6

Confidence Intervals

	Estimate	LowerCI	UpperCI
(Intercept)	2.693190069	-0.9686476256	6.355027763
Ln $Trade_{ij}$	0.285529494	0.2685436230	0.302515365
Ln $GDPPC_j$	0.184437091	0.1320035407	0.236870642
Ln $GDPPC_i$	0.306853660	0.2606375381	0.353069781
Ln POP_j	-0.003713246	-0.1713877872	0.163961295
Ln POP_i	0.191142444	0.0225924184	0.359692469
Ln PPP_{ij}	-0.108312441	-0.1355546685	-0.081070213
Ln $Dist_{ij}$	-1.083815477	-1.1321118122	-1.035519141
$Lang_{ij}$	0.832292799	0.7441951885	0.920390409
$Border_{ij}$	1.116114661	0.9251400326	1.307089289
$Colony_{ij}$	0.596346251	0.3797124599	0.812980043
CU_{ij}	0.212288639	0.0045069814	0.420070297
Event	0.027348131	0.0012520930	0.053444168
Event (t+1)	-0.060808654	-0.0884901234	-0.033127185
Event (t+2)	-0.078980292	-0.1088017254	-0.049158858
Event (t+3)	-0.063605168	-0.0965298590	-0.030680476
Event (t-1)	0.041088724	0.0117413723	0.070436075
Event (t-2)	0.024862796	-0.0014072943	0.051132887
Event (t-3)	0.028803287	-0.0009791786	0.058585753

Table D7

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	2.691807094	-0.966253471	6.3498676596	2.7234814836	-0.9332416673	6.380204634
Ln Trade$_{ij}$	0.285548460	0.268566191	0.3025307283	0.2854906805	0.2685078166	0.302473544
Ln GDPPC$_j$	0.185071342	0.132653994	0.2374886894	0.1855142969	0.1331028929	0.237925701
Ln GDPPC$_i$	0.302032186	0.255994488	0.3480698848	0.3030444091	0.2570362176	0.349052601
Ln POP$_j$	-0.003385849	-0.170964799	0.1641931004	-0.0029082720	-0.1704870439	0.164670500
Ln POP$_i$	0.193324489	0.024863215	0.3617857619	0.1898892065	0.0214891119	0.358289301
Ln PPP$_{ij}$	-0.107614601	-0.134824622	-0.0804045802	-0.1072149885	-0.1344382634	-0.079991713
Ln Dist$_{ij}$	-1.083768669	-1.132061684	-1.0354756533	-1.0838976442	-1.1321910295	-1.035604259
Lang$_{ij}$	0.832018474	0.743929913	0.9201070342	0.8322214836	0.7441231668	0.920319800
Border$_{ij}$	1.115789956	0.924825834	1.3067540789	1.1160689844	0.9251016805	1.307036288
Colony$_{ij}$	0.595691230	0.379058967	0.8123234931	0.5963034032	0.3796763215	0.812930485
CU$_{ij}$	0.211959976	0.004207082	0.4197128708	0.2120823555	0.0043162451	0.419848466
Participant	0.163681651	0.117833263	0.2095300395			
Non Participant	-0.038052607	-0.082299382	0.0061941687			
Peak Season				-0.0949101009	-0.1271753604	-0.062644842
Off Season				0.1099008960	0.0738173325	0.145984460

Table D8

Confidence Intervals

	Estimate	LowerCI	UpperCI
(Intercept)	2.728706053	-0.929196211	6.3866083163
Ln Trade$_{ij}$	0.285520749	0.268537892	0.3025036065
Ln GDPPC$_j$	0.184795411	0.132361351	0.2372294710
Ln GDPPC$_i$	0.303157807	0.257104383	0.3492112309
Ln POP$_j$	-0.005698656	-0.173414906	0.1620175939
Ln POP$_i$	0.192952494	0.024459016	0.3614459714
Ln PPP$_{ij}$	-0.107277229	-0.134496139	-0.0800583188
Ln Dist$_{ij}$	-1.083865643	-1.132158516	-1.0355727698
Lang$_{ij}$	0.832241257	0.744143552	0.9203389619
Border$_{ij}$	1.116028258	0.925057200	1.3069993164
Colony$_{ij}$	0.596335503	0.379702219	0.8129687869
CU$_{ij}$	0.212012813	0.004251492	0.4197741346
Bid Host	0.018992826	-0.006653667	0.0446393193
Bid Candidate	0.016476655	-0.004606076	0.0375593860

Table D9

Confidence Intervals

	Estimate	LowerCI	UpperCI
(Intercept)	-1.435626342	-5.196635746	2.32538300
Ln Trade$_{ij}$	0.285533116	0.268496903	0.30256930
Ln GDPPC$_{j}$	0.187701747	0.135042497	0.24036100
Ln GDPPC$_{i}$	0.231644741	0.185617253	0.27767220
Ln POP$_{j}$	0.02482759	-0.142229655	0.19188480
Ln POP$_{i}$	0.415379488	0.242026165	0.58873280
Ln PPP$_{ij}$	-0.101990491	-0.127606737	-0.07637425
Ln Dist$_{ij}$	-1.083728823	-1.131711114	-1.03574700
Lang$_{ij}$	0.832427516	0.743950007	0.92090500
Border$_{ij}$	1.118665249	0.926613286	1.31071700
Colony$_{ij}$	0.594426266	0.377167593	0.81168490
CU$_{ij}$	0.210092519	0.002874626	0.41731040
Ln ORK$_{j}$	0.015781228	-0.142060102	0.17362260
Ln ORK$_{i}$	1.083858936	0.929170153	1.23854800
Event	0.037781946	0.016304039	0.05925985

Table D10

Confidence Intervals

	Estimate	LowerCI	UpperCI
(Intercept)	-1.929315826	-5.6629159465	1.804284295
Ln Trade$_{ij}$	0.285807600	0.2687819645	0.302833235
Ln GDPPC$_j$	0.198546213	0.1469063189	0.250186108
Ln GDPPC$_i$	0.216055273	0.1698371704	0.262273376
Ln POP$_j$	0.028057009	-0.1380064128	0.194120431
Ln POP$_i$	0.447070133	0.2750106854	0.619129581
Ln PPP$_{ij}$	-0.080670723	-0.1051405283	-0.056200918
Ln Dist$_{ij}$	-1.083438814	-1.1313921698	-1.035485458
Lang$_{ij}$	0.831774459	0.7432989200	0.920249999
Border$_{ij}$	1.118127569	0.9260918262	1.310163311
Colony$_{ij}$	0.594848358	0.3775657099	0.812131006
CU$_{ij}$	0.209265062	0.0020329507	0.416497173
Ln Corruption$_j$	-0.036181396	-0.1594248210	0.087062029
Event	0.032659030	0.0113212788	0.053996781
Ln Govt Effec$_j$	0.117789831	-0.0335846267	0.269164289
Ln Rgltry Qlty$_j$	-0.107045842	-0.2296519839	0.015560299
Ln Politcl Stab$_j$	0.027837771	-0.0134205016	0.069096043
Ln Rul of Law$_j$	0.010715546	-0.1412540661	0.162685159
Ln Corruption$_i$	-0.177130862	-0.2940472086	-0.060214516
Ln Govt Effec$_i$	1.183819490	1.0440071867	1.323631793
Ln Politcl Stab$_i$	0.184505631	0.1325458038	0.236465459
Ln Rgltry Qlty$_i$	-0.133227532	-0.2424388993	-0.024016165
Ln Rul of Law$_i$	0.132902948	-0.0055621876	0.271368084

Table D11

Confidence Intervals

	Estimate	LowerCI	UpperCI
(Intercept)	-1.443360707	-5.204501528	2.3177801133
Ln $Trade_{ijt}$	0.285541392	0.268505302	0.3025774812
Ln $GDPPC_{jt}$	0.187637968	0.134974520	0.2403014162
Ln $GDPPC_{it}$	0.232128828	0.186084611	0.2781730447
Ln POP_{jt}	0.024762300	-0.142341919	0.1918665200
Ln POP_{it}	0.415796238	0.242431601	0.5891608746
Ln PPP_{ij}	-0.102057689	-0.127675615	-0.0764397623
Ln $Dist_{ij}$	-1.083718613	-1.131700952	-1.0357362751
$Lang_{ij}$	0.832412658	0.743935668	0.9208896479
$Border_{ij}$	1.118675921	0.926623076	1.3107287661
$Colony_{ij}$	0.594426293	0.377167636	0.8116849510
CU_{ij}	0.210110888	0.002893532	0.4173282436
Ln ORK_{jt}	0.015770220	-0.142066757	0.1736071978
Ln ORK_{it}	1.082502820	0.927719873	1.2372857665
Bid Host	0.020229221	-0.005728836	0.0461872782
Bid Candidate	0.004715469	-0.016488705	0.0259196417

Table D12

Confidence Intervals

	Estimate	LowerCI	UpperCI	Estimate	LowerCI	UpperCI
(Intercept)	-1.459655233	-5.219348487	2.300038022	-1.482305658	-5.243111275	2.278499958
Ln $Trade_{ijt}$	0.285510376	0.268474229	0.302546522	0.285562674	0.268527348	0.302598001
Ln $GDPPC_{jt}$	0.188226469	0.135590816	0.240862122	0.187759180	0.135113190	0.240405169
Ln $GDPPC_{it}$	0.231973607	0.185970362	0.277976851	0.230954135	0.184927015	0.276981256
Ln POP_{jt}	0.027400537	-0.139563200	0.194364274	0.026849914	-0.140114021	0.193813848
Ln POP_{it}	0.413683775	0.240379766	0.586987784	0.416710965	0.243356452	0.590065478
Ln PPP_{ij}	-0.102013253	-0.127631042	-0.076395464	-0.102425601	-0.128032889	-0.076818312
Ln $Dist_{ij}$	-1.083757620	-1.131740299	-1.035774941	-1.083632893	-1.131615535	-1.035650250
$Lang_{ij}$	0.832371784	0.743893720	0.920849849	0.832197062	0.743728904	0.920665219
$Border_{ij}$	1.118745165	0.926696299	1.310794031	1.118533115	0.926491728	1.310574503
$Colony_{ij}$	0.594381892	0.377129695	0.811634088	0.593805892	0.376546925	0.811064860
CU_{ij}	0.210185447	0.002962086	0.417408808	0.210049875	0.002836882	0.417262868
Ln ORK_{jt}	0.016206386	-0.141542761	0.173955533	0.016155509	-0.141667114	0.173978132
Ln ORK_{it}	1.083771370	0.929106979	1.238435760	1.082257861	0.927580950	1.236934771
Peak Season	-0.096725809	-0.129184877	-0.064266740			
Off Season	0.110859974	0.074198291	0.147521657			
Event Part				0.166155145	0.120935142	0.211375148
Event No Part				-0.030901829	-0.075949227	0.014145568

VITA

Abdul Sami Stanekzai
Department of Engineering Management and Systems Engineering
2101 Engineering Systems Building
Frank Batten College of Engineering and Technology
Old Dominion University
Norfolk, Virginia 23529

Abdul Sami Stanekzai completed his Bachelor of Science degree in Mechanical Engineering at Kabul University of Kabul, Afghanistan in December 2011. Mr. Stanekzai is the recipient of the Fulbright Scholarship from U.S. Department of State in 2013, through which he started his current Master's degree program in August 2014. His current research interests are in Operations Research and optimization algorithms. Before joining Old Dominion University for his graduate studies, Mr. Stanekzai worked with some of the well-known U.S Army Corps of Engineers' (USACE) prime contractors on projects inside Afghanistan and other middle eastern countries including Qatar, and Bahrain. From 2011 to 2014 he held different technical and managerial positions ranging from Quality Control (QC) Engineer to Deputy QC Manger and Project manager when worked with USACE. In 2011, Mr. Stanekzai also started his own startup engineering firm, through which he successfully handed over several medium scale projects to U.S. Army support teams in Afghanistan as well as the Government Islamic Republic of Afghanistan.