

# The Impact of Length of Stay on Visitor Spending During The 2034 World Cup: An Instrumental Variable

Walaa Ahmed Hamdi \*

Department of Mathematics and Statistics, University of Jeddah, Jeddah, Saudi Arabia

\*Corresponding author E-mail: [whamdi@uj.edu.sa](mailto:whamdi@uj.edu.sa)

Received: March 23, 2026, Accepted: April 28, 2026, Published: May 10, 2026

## Abstract

This study examines the causal effect of stay length on visitor spending during the 2034 FIFA World Cup in Saudi Arabia. Using a large-scale simulated dataset informed by FIFA projections, the analysis employs Two-Stage Least Squares (2SLS) regression to address potential endogeneity and compare with Ordinary Least Squares (OLS). Package tour participation serves as the instrument, and the age group is included as a control. The results indicate a strong, statistically significant positive effect of length of stay on visitor spending for both methods. The instrumental variable estimates further confirm the robustness of this relationship. These findings underscore the economic importance of encouraging longer visitor stays and provide methodological insights for analyzing the impacts of mega-events under data constraints.

**Keywords:** Causal effect; FIFA World Cup; Instrumental Variable; Length of Stay; Tourism Economics; Two-Stage Least Squares.

## 1. Introduction

Saudi Arabia is a powerful country in both tourism and the economy at present. It has successfully hosted diverse events, including sports, entertainment, cultural, and business conferences. Moreover, Saudi Arabia has welcomed many major international sporting events across its cities, such as Formula 1 races, the Spanish and Italian Super Cups, and the Esports World Cup. The inflow of tourists into Saudi Arabia is steadily increasing toward its ambitious goal of attracting up to 150 million tourists annually by 2030 [1]. Saudi Arabia is positioning itself as a welcoming tourist destination with activities for all visitors.

Global sporting events are widely regarded as important drivers of tourism development and broader economic activity, in line with the objectives of Saudi Vision 2030. In this context, hosting the 2034 FIFA World Cup is expected to further support growth in tourism and related sectors. However, understanding the factors that influence visitor spending remains essential for evaluating the economic impact of such events; see [2].

A key determinant of visitor spending for tourism is the length of stay, which directly influences demand for accommodations, transportation, entertainment, and other services. Therefore, longer stays tend to yield higher expenditures on goods and services, as demonstrated by a tourism study conducted in Portugal by [3]. Estimating the causal impact of longer stays on spending is complicated by potential endogeneity, according to [4].

This study will examine the relationship between length of stay and visitor spending in the context of the 2034 FIFA World Cup. As the event has not yet taken place, the analysis relies on a simulated dataset, which provides a controlled framework for studying this relationship. A package tour indicator is used as an instrumental variable to address potential endogeneity in length of stay, while Age Group is included as a control variable to capture demographic differences in travel behavior. The study applies a Two-Stage Least Squares (2SLS) estimation approach to analyze the relationship between stay length and spending.

The finding will contribute to the tourism and economic literature by providing a structured, simulation-based analysis of visitor spending in the context of a major international sporting event. The study is intended as a methodological illustration rather than a source of direct policy conclusions, with implications that are consistent with the objectives of Saudi Vision 2030.

The article is organized as follows: Section 2 reviews the relevant literature, Section 3 presents the data and methodology, Section 4 reports the estimation results and provides their interpretation, and Section 5 concludes the study.

## 2. Literature Review

Many studies have emphasized the critical role of length of stay (LoS) in determining tourist spending, with longer stays generally associated with higher expenditures on accommodation, food, and local services; see [3], [5]. However, the strength of this relationship varies across contexts and depends on model specification and unobserved heterogeneity in tourist behavior.

A key challenge in this literature is endogeneity. Length of stay is likely to be jointly determined with spending due to unobserved factors such as income, travel preferences, and trip motivation, as well as potential reverse causality. As a result, ordinary least squares (OLS) estimates may suffer from bias and inconsistency; see [4].

To address these limitations, instrumental variable (IV) techniques, particularly two-stage least squares (2SLS), are widely used to obtain consistent estimates under endogeneity; see [6–7]. Within this framework, package tours have been proposed as a plausible instrument, as they influence trip duration through pre-arranged itineraries while being assumed not to directly affect individual spending decisions; see [4]. However, this exclusion restriction is context-dependent and may not fully hold in practice due to bundled pricing and consumption structures associated with package travel.

Despite a substantial body of literature on tourism expenditure and length of stay, existing studies remain largely based on observational data and case-specific contexts. There is limited examination of the relationship between LoS and spending in the context of large-scale international sporting events. This study addresses this gap by providing a simulation-based instrumental variables (IV) framework that allows transparent control over endogeneity and identification assumptions, applied to the future 2034 FIFA World Cup in Saudi Arabia.

### 3. Data and Methodology

#### 3.1. Simulated data

This analysis will examine the impact of the length of stay on visitors' spending during the FIFA World Cup in Saudi Arabia. Hosting international events offers certain benefits, such as increased spending and tourism, which in turn boost Gross Domestic Product (GDP); see [6]. Using simulated data provides a controlled environment in which the causal impact of LoS can be assessed while addressing potential endogeneity concerns through an instrumental variable (IV) approach. It is important to note that the purpose of the simulation is methodological, aiming to illustrate the application of IV estimation under realistic behavioral assumptions rather than to provide direct empirical policy conclusions.

In this study, we want to capture the diversity of tourist behavior better. The simulated dataset incorporates key visitor characteristics such as Age Group and a Package Tour. Age is categorized into three groups: Young, Middle-aged, and Senior, while travel type is represented by a binary indicator for participation in either a package tour or an independent trip. These variables are supported by empirical findings; previous research has shown that both age and package tour are positively associated with longer lengths of stay; [8]. Incorporating these characteristics allows the model to reflect heterogeneity in spending patterns more realistically, recognizing that travel decisions and economic contributions can differ across demographic groups and travel arrangements.

The modeling framework facilitates a structured understanding of the economic impact of tourism in the context of hosting a major international sporting event, specifically, Saudi Arabia will be hosting the FIFA World Cup 2034, which will feature a 48-team format in one country. According to the Saudi Arabia FIFA World Cup 2034 bid documents [9], which projects overall attendance in the range of approximately 5 to 7 million visitors, the simulation assumes a total of 6,000,000 visitors to analyze the effect of length of stay on tourist spending in Saudi Arabia. The key variables used in the analysis are defined as follows:

- **Length of Stay (LoS):** Simulated as a discrete variable bounded between 2 and 10 days to reflect realistic tourist behavior. The variable is generated from an underlying latent continuous process and then rounded and truncated. LoS is treated as an endogenous outcome, influenced by Package Tour participation, Age Group, and an unobserved error component. The data-generating process combines deterministic effects with stochastic variation from normally distributed shocks, producing realistic dispersion in travel duration. Endogeneity arises because the unobserved  $u$  factor affects both Length of Stay and Spending, inducing correlation between LoS and the error term in the structural spending.
- **Spending per Visitor:** Generated from an exponential transformation of a heteroskedastic log-linear structural model with a quadratic term in Length of Stay to capture diminishing marginal returns. The model is specified in logarithmic form and includes heteroskedastic Gaussian disturbances whose variance increases with Length of Stay. In addition, a shared unobserved component enters both the Length of Stay and Spending equations, inducing endogeneity. After exponentiation, spending becomes strictly positive and right-skewed, producing a distribution that generalizes the log-normal case and reflects multiplicative heterogeneity in spending behavior. Age Group effects are incorporated through additive shifts in the log-scale specification, capturing systematic differences in expenditure across demographic categories. This structure produces a nonlinear, heteroskedastic, and highly skewed distribution driven by both functional form and stochastic variation.
- **Package Tour:** Package Tour is a binary treatment indicator of travel type (1 = package tourist; 0 = independent traveler), generated from a Bernoulli distribution with probability given by a logistic transformation of an exogenous standard normal variable  $z$ . This specification ensures variation in generating a positive correlation with the Length of Stay. Package Tour is used as an instrumental variable due to its relevance in explaining variation in travel duration, while remaining exogenous with respect to the unobserved error in the spending equation.
- **Age Group:** Age Group is a categorical variable dividing visitors into Young, Middle-aged, and Senior categories. It is generated using fixed probabilities (40%, 40%, and 20%, respectively), reflecting a plausible demographic structure of international visitors. Age Group captures heterogeneity in both travel behavior and spending patterns and enters both the LoS and Spending equations through deterministic additive effects.

The data-generating process is characterized by:

- a latent continuous Length of Stay process discretized and constrained within [2,10],
- a nonlinear log-linear spending equation with a quadratic LoS term,
- heteroskedastic normal disturbances in the log-spending equation with variance increasing in Length of Stay,
- and multiplicative heterogeneity in spending behavior induced by exponentiation of the log-spending equation.

Together, these features generate a realistic tourism dataset suitable for a two-stage least squares, where Package Tour is used as an instrument for Length of Stay. The effect of package tours on Length of Stay is captured through the assignment mechanism, generating a positive association consistent with the idea that organized travel arrangements influence trip duration. Previous study suggests that participation in package tours generally leads to an increase in length of stay by an average of 1 to 2 days, see [10].

### 3.1.1. Endogenous regressor

In the estimation of the causal impact of length of stay (LoS) on spending per visitor, we employ a two-stage least squares (2SLS) regression approach, which addresses potential endogeneity in LoS. Endogeneity may arise from unobserved factors, such as individual preferences or budget constraints, that influence both the duration of stay and spending behavior. To address this issue, Package Tour is used as an instrumental variable. This method is well-established for correcting endogeneity in observational data, particularly in economic applications; see [7]. The Age Group is included as a control variable in both stages of the analysis to account for demographic variation in travel behavior. For example, the older tourists may have more flexible schedules and different consumption patterns, affecting both LoS and spending decisions, but the FIFA World Cup targets all Age Groups around the world. In this model, including Age Group as a control variable helps improve the credibility of the estimated effect on visitor spending.

The first-stage regression is specified as:

$$\text{Length\_of\_Stay}_i = \pi_0 + \pi_1 \text{Package\_Tour}_i + \pi_2 \text{Age\_Group}_i + \epsilon_i \quad (1)$$

Where:

- $\pi_0$  is the intercept, representing the baseline length of stay for a visitor not participating in a package tour and belonging to the reference Age Group.
- $\pi_1$  represents the average effect of participating in a package tour on Length of Stay. Based on the simulation design, package tours are associated with an increase in the expected duration of stay. This effect is introduced with stochastic variation rather than a fixed deterministic shift, allowing for more realistic behavioral patterns.
- $\pi_2$  reflects the effect of Age Group on Length of Stay, where older age categories are associated with longer stays, consistent with empirical findings reported in [8].
- $\epsilon_i$  is a composite disturbance term containing a shared unobserved component ( $u$ ) that induces endogeneity between Length of Stay and Spending, together with an idiosyncratic normally distributed shock.

## 3.2. Regression models

In this section, we consider the econometric models used to assess the effect of Length of Stay (LoS) on spending per visitor during the 2034 FIFA World Cup using the simulation data. We estimate both an Ordinary Least Squares (OLS) regression and a Two-Stage Least Squares (2SLS) regression to correct for potential endogeneity in LoS.

### 3.2.1. Ordinary least squares (OLS) model

The Ordinary Least Squares (OLS) model serves as a baseline to quantify the relationship between the length of stay in Saudi Arabia and the total spending. Age Group is included as a control variable to account for demographic differences in spending behavior. The OLS regression model is expressed as follows:

$$\text{Spending}_i = \beta_0 + \beta_1 \text{Length\_of\_Stay}_i + \beta_2 \text{Age\_Group}_i + \epsilon_i \quad (2)$$

Where:

- $\text{Spending}_i$ : Total expenditure per visitor, generated from a heteroskedastic log-linear model that ensures a strictly positive and right-skewed distribution,
- $\text{Length\_of\_Stay}_i$ : Number of days the visitor remains in Saudi Arabia,
- $\beta_1$ : Marginal effect of one additional day on total spending,
- $\text{Age\_Group}_i$ : Categorical variable (Young, Middle-aged, Senior) used as a control,
- $\beta_2$ : Coefficients for Age Group categories,
- $\epsilon_i$ : Error term capturing unobserved factors affecting spending.

The OLS model gives a simple estimate of the relationship between the Length of Stay and spending. However, if LoS is endogenous, it may produce biased and inconsistent estimates. In this simulation, endogeneity is controlled through the data-generating process, but IV estimation is still applied to illustrate methodological application.

The Package Tour is used as an instrumental variable because it is strongly associated with the Length of Stay. Within the simulation framework, the exclusion restriction is imposed by construction, such that the effect of Package Tour on spending operates primarily through its influence on Length of Stay. Nevertheless, in real-world applications, this assumption may be more difficult to justify, as package tours can also directly affect spending through how services are combined and priced.

### 3.2.2. Two-stage least squares (2SLS) / instrumental variables model

A Two-Stage Least Squares (2SLS) regression is implemented using the Package Tour as an instrument for LoS. Age Group is included as a control variable in both stages (not instrumented). In the second stage, the predicted values of LoS from the first stage are used to estimate the relationship between stay duration and total spending:

$$\text{Spending}_i = \beta_0 + \beta_1 \widehat{\text{Length\_of\_Stay}}_i + \beta_2 \text{Age\_Group}_i + \epsilon_i \quad (3)$$

In this simulation framework, Package Tour is constructed as a valid instrument for Length of Stay through a logistic function of an exogenous standard normal variable. The exclusion restriction is imposed by design, since ensure that Package Tour affects spending only through its impact on Length of Stay, while controlling for Age Group. The 2SLS framework provides an estimate of  $\beta_1$ , representing the association between extended tourist stays and spending within the simulation. However, in real-world applications, this assumption may be violated if package tours include combined services that directly affect expenditure levels.

## 4. Estimation Results and Interpretation

The results of analysis examining the relationship between Length of Stay and Spending per Visitor during the FIFA World Cup 2034, using a Two-Stage Least Squares (2SLS) approach, address potential endogeneity in Length of Stay by using Package Tour as an instrument, while keeping Age Group as a control variable in both stages. All estimations were conducted using R software, ensuring reproducibility and robust statistical output.

### 4.1. First-stage regression

The first-stage regression model investigates the predictive power of the instrument on the endogenous variable, Length of Stay. The regression results show that the instrument has a statistically significant and economically meaningful effect on Length of Stay:

Package Tour has a statistically significant positive effect on Length of Stay, where package tourists stay approximately 0.40 days longer than independent travelers ( $t = 384.9$ ,  $p < 0.001$ ).

Age Group is statistically significant, indicating systematic differences in Length of Stay across demographic groups:

- Middle-aged visitors stay longer than young visitors by approximately 0.43 days.
- Senior visitors stay longer by approximately 0.89 days.

The first-stage results were influenced by both the large sample size and the relatively low noise structure in the simulation design.

The first-stage F-statistic is 193,800, which is far above the conventional threshold of 10, confirming that Package Tour is a strong instrument and there is no evidence of weak instrument bias. The R-squared of approximately 0.088 indicates modest explanatory power, which is expected given the stochastic structure of the simulated data.

### 4.2. Second-stage regression (instrumental variable model)

The second-stage regression estimates the causal effect of Length of Stay on Spending per Visitor, using Package tour as an instrument. The IV results show:

- Each additional day of stay causes an increase of approximately 344.62 SAR in spending per visitor ( $t = 286.9$ ,  $p < 0.001$ ).
- The IV estimate is lower than the OLS estimate (508.28 SAR), suggesting that OLS is upward-biased due to the positive correlation between Length of Stay and unobserved factors affecting spending.
- Age Group has statistically significant, but with a relatively small economic magnitude, indicating moderate demographic differences in spending behavior.

The difference between OLS and IV suggests endogeneity, and IV is the preferred estimator. The statistical significance of the estimates is also amplified by the large sample size and controlled variance in the simulated data.

### 4.3. Model diagnostics

#### 4.3.1. Weak instrument test

The first-stage results indicate a strong relationship between Package Tour (instrument) and Length of Stay, with an F-statistic of 148,141. This value is far above the conventional threshold of 10, providing strong evidence against weak instrument concerns. The magnitude of the F-statistic is partly influenced by the large sample size and the relatively low-variance simulation structure. The overall first-stage model F-statistic is 193,800, further confirming strong instrument relevance.

#### 4.3.2. Endogeneity test (Wu–Hausman)

The Wu–Hausman test is used to examine whether Length of Stay (LoS) is endogenous in the model. The test results show a statistic of 5,748 with  $p < 0.001$ . The strong rejection of exogeneity reflects the simulated data-generating process, where Length of Stay is explicitly correlated with unobserved heterogeneity.

#### 4.3.3. Heteroskedasticity test

The Breusch–Pagan test results show a test statistic of  $BP = 190,694$  with a p-value less than 0.001, indicating rejection of the null hypothesis of homoskedasticity. This provides evidence of heteroskedasticity in the residuals and is consistent with the simulation design, where variance increases with Length of Stay.

#### 4.3.4. Robustness checks

The results show a highly consistent relationship between Length of Stay and Spending per Visitor. OLS and IV estimates differ substantially, supporting the presence of endogeneity in the OLS specification. The instrument (Package Tour) is strong and statistically valid. The diagnostic tests collectively confirm endogeneity in Length of Stay and justify the use of IV estimation. Heteroskedasticity is also present, consistent with the underlying data-generating process.

## 5. Conclusion

This study examines the relationship between Length of Stay and visitor spending of the 2034 FIFA World Cup in Saudi Arabia using a large-scale simulated dataset. The analysis employs Two-Stage Least Squares (2SLS) regression to account for potential endogeneity in Length of Stay, with Package Tour serving as an instrument and Age Group included as a control variable.

The results show that a positive and statistically significant relationship between Length of Stay and Spending per Visitor. However, the instrument variable (IV) estimate is consistently lower than the OLS estimates, suggesting that the baseline OLS model is upward-biased, and there is unobserved heterogeneity embedded in the simulated data-generating process. The IV estimate provides a more credible identification strategy for isolating the causal effect of Length of Stay within the simulation framework.

Diagnostic tests further confirm the relevance and strength of the instrument, as well as the presence of endogeneity and heteroskedasticity in the model. These results highlight the importance of instrumental variable techniques when addressing omitted variable bias and correlated unobservables in structural estimation.

It is important to emphasize that the findings of this study are derived entirely from simulated data. Therefore, the results are not for empirical policy estimates of tourism planning related to the FIFA World Cup 2034. Moreover, the main contribution of this study is methodological when demonstrating how IV estimation can be implemented under controlled conditions to recover structural relationships in the presence of endogeneity.

Future studies can examine the causal effect using IV by applying it to real data from FIFA 2034 or similar real-world data. Furthermore, future research may extend this framework by incorporating real-world tourism datasets or alternative sources of exogenous variation, such as travel costs, transport accessibility, visa requirements, or booking constraints, to enhance external validity and empirical applicability.

## Acknowledgement

The authors would like to express their sincere gratitude to the reviewers for their valuable suggestions, which have significantly improved the quality of this paper.

## References

- [1] Saudi Vision 2030: Transforming Tourism for the Future. Kingdom of Saudi Arabia, available online: [www.vision2030.gov.sa](http://www.vision2030.gov.sa), last visit: 7 Aug. 2025.
- [2] Loftsgaarden, B., The Economic Impact of Major Events, *Economic Development Journal*, Vol. 16, No.4, (2017), pp.44–50, available online: [EDJ\\_17\\_Fall\\_Loftsgaarden.pdf](http://EDJ_17_Fall_Loftsgaarden.pdf).
- [3] Barros, C., Butler, R., and Correia, A., The Length of Stay of Golf Tourists in the Algarve: A Survival Analysis, *Tourism Management*, Vol.31, No.1, (2010), pp.13–21, available online: <https://doi.org/10.1016/j.tourman.2009.02.010>.
- [4] Thrane, C., On the Relationship between Length of Stay and Total Trip Expenditures: A Case Study of Instrumental Variable (IV) Regression Analysis, *Tourism Economics*, Vol.21, No. 2, (2015), pp.357-367, available online: <https://doi.org/10.5367/te.2014.0357>.
- [5] Wang, Y., and Davidson, M., A review of micro-analyses of tourist expenditure, *Current Issues in Tourism*, Vol.13, No.6, (2010), pp.507–524, available online: <https://doi.org/10.1080/13683500903406359>.
- [6] Assaf, A.G., and Tsionas, M., Correcting for endogeneity in hospitality and tourism research, *International Journal of Contemporary Hospitality Management*, Vol.32, No.8, (2020), pp.2657–2675, available online: <https://doi.org/10.1108/IJCHM-02-2020-0153>.
- [7] Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data (2nd ed.)*, Cambridge: MIT Press, (2010).
- [8] Oklevik O., Kwiatkowski G., Malchrowicz-Moško E., Ossowska L., and Janiszewska D., Determinants of tourists' length of stay, *PLoS ONE* Vol.16, No.12, (2021), available online: <https://doi.org/10.1371/journal.pone.0259709>.
- [9] Saudi Arabia FIFA World Cup™ 2034 Bid. (n.d.), available online: <https://saudi2034.com.sa/>, last visit: 7 Aug. 2025.
- [10] Chen, Y., Schuckert, M., Song, H. and Chon, K.. Why Can Package Tours Hurt Tourists? Evidence from China's Tourism Demand in Hong Kong, *Journal of Travel Research*, Vol.55, (2015), pp.427-439, available online: <https://doi.org/10.1177/0047287515612597>.