What Happens in Vegas Stays in Vegas? Firsthand Experience and EDGAR Search Activity in Las Vegas Casino Hotels

Ryan Flugum
Department of Finance
College of Business
University of Northern Iowa
Cedar Falls, IA 50613
ryan.flugum@uni.edu

Choonsik Lee
College of Business
University of Rhode Island
Kingston, RI 02881
choonsiklee@uri.edu

Matthew E. Souther
Department of Finance
Darla Moore School of Business
University of South Carolina
Columbia, SC 29208
matthew.souther@moore.sc.edu

I. Data Sets

Vegas_data_fm.csv: This is a pseudo-dataset that has the same data structure as the actual dataset used to produce the results of Tables 1 – 5 and Tables A2 and A3 of the appendix, as well as Figures 2 and 4. Permno, Gvkey, and all IP-related data have been anonymized or randomized within the same range of actual values in the file due to licensing agreements. This dataset is used in both the Table_Code_FM.do and Table_Code_FM.sas files.

The variable definitions in this dataset are as follows:

Permno = CRSP identifier (anonymized)
gvkey = Compustat identifier (anonymized)
date = Monthly observation date
ffi48 = Fama-French 48 industry classification
sic4 = Four-digit SIC code
mtime = Monthly indexing variable that is year x 12+ month
vegas_access = Count variable of the number of EDGAR access events taking place in a Las Vegas casino hotel in the current month for the firm's SEC documents. We exclude all access events labeled as web crawler in the SEC Log File. (randomized)
vegas_ind = An indicator variable that is one if the firm has at least one access event taking place in a Las Vegas casino hotel in the current month, and zero otherwise. (randomized)
vegas_firm = An indicator that is one if the firm has a four-digit sic code that is between 4100-4100 (Transit and passenger trans), 4110-4119 (Local passenger trans), 4120-4121 (Taxicabs), 4140-4142 (Bus charter), 4190-4199 (Misc transit and passenger transportation), 4240-4249 (Transportation), 4500-4599 (Air transportation), 4700-4700 (Transportation services), 4720-4729 (Travel agencies), 5800-5819 (Retail - eating places), 5820-5829 (Restaurants, hotels, motels), 5890-5899 (Eating and drinking places), 7000-7000 (Hotels, other lodging places), 7010-7019 (Hotels motels), 7040-7049 (Membership hotels and lodging), 7800-7829 (Services - motion picture production and distribution), 7830-7833 (Services - motion picture theatres), 7840-7841 Services - video rental), 7900-7900 (Services - amusement and recreation), 7920-7929 (Services - bands, entertainers),

7940-7949 (Services - professional sports), 7980-7980 (Amusement and recreation services), or 7990-7999 (Services - misc entertainment). (randomized)
vegas_firm_interaction = Interaction term of vegas_ind and vegas_firm (randomized)
vegas_ba = An indicator that is one if the firm is a Vegas firm based on SIC code AND the firm also has a Las Vegas business or mailing address in their most recent 10K filings with the SEC.
f6 = This is the aggregate number of 8-K filings made by the firm over the prior 6 months, following Zhao (2017). In our multivariate tests, we use the natural log of one plus the number of 8-K filings during this time period. (randomized)
log_f6 = The natural log of one plus f6. (randomized)
allfilings = The natural log of total filings the firm makes with the SEC in the respective month. (randomized)
log_allfilings = The natural log of total filings the firm makes with the SEC in the respective month. (randomized)
ra = This is the total EDGAR access events for the specific CIK in the respective month, divided by the sum of all EDGAR access events in the respective month. (randomized)
gs = We use the average of search volume for a firm's ticker, ticker "price," and ticker "stock," which we obtain via Google Trends at a monthly frequency (Chen, Schmidt, and Wang, 2021). The Google search index ranges from 0 to 100, giving a phrase a value of 100 if it is the month with the most search of the respective phrase over the designated time period. We consider general web searches in the Finance category of all three phrases simultaneously. The final measure we use in our multivariate tests is the average of the search index values across the three phrases we consider for each firm over the time period 2004 to 2022. (randomized)
earn = An indicator variable that is one if the firm announces earnings in the respective month and zero otherwise.
lmcap = The market capitalization of the firm as of the respective month. We obtain the firm's share price and shares outstanding from the CRSP database. In our multivariate test, we winsorize this variable at the 1% level and take natural log throughout our multivariate tests.
lmb = The firm's book-to-market ratio computed following Fama and French (2008) and using a firm's most recent annual financial statements. We winsorize this variable at the 1% level and take natural log throughout our multivariate tests.
roa_w = The firm's return on assets is computed using a firm's most recent financial statement data in Compustat. Specifically, we compute return on assets as a firm's net income divided by the average of a firm's total assets over the two most recent years. We winsorize this variable at the 1% level.
lev_w = The firm's book leverage which we compute using a firm's most recent annual financial statement data in Compustat. We winsorize this variable at the 1% level.
ret12_w = The firm's cumulative return over the prior twelve months and winsorized at the 1% level.
std12_w = The firm's monthly standard deviation of returns over the prior twelve months and winsorized at the 1% level.
lag_idiovol_w = Following Ang, Hodrick, Xing, and Zhang (2006), we estimate idiosyncratic volatility as the standard deviation of each firm's daily residuals from a Carhart (1997) four-factor model estimated using the prior six months of daily returns.

io_w = The firm's institutional ownership percentage as of the most recent quarter and winsorized at the 1% level.
ret1 = The return of the firm in the subsequent month.
ret3 = The return of the firm in the subsequent 3 months.
ret6 = The return of the firm in the subsequent 6 months.
ret121 = The return of the firm in the subsequent 12 months.
Bhar1m = The Buy and Hold Abnormal Return of the firm in the subsequent month, using the CRSP Value Weighted index as the benchmark return.
Bhar3m = The Buy and Hold Abnormal Return of the firm in the subsequent 3 months, using the CRSP Value Weighted index as the benchmark return.
Bhar6m = The Buy and Hold Abnormal Return of the firm in the subsequent 6 months, using the CRSP Value Weighted index as the benchmark return.
Bhar12m = The Buy and Hold Abnormal Return of the firm in the subsequent 12 months, using the CRSP Value Weighted index as the benchmark return.

Ipaccess_fm.csv: This is a pseudo-dataset that has the same data structure as the actual dataset containing the specific IP access events. We use the IP access events to compute our access-related variables (i.e., vegas_access, vegas_ind). Permno, Gvkey, and all IP-related data have been anonymized or randomized within the same range of actual values in the file due to licensing agreements. This dataset is used in the Table_Code_FM.sas files.

The variable definitions in this dataset are as follows:

Permno = CRSP identifier (anonymized)
gvkey = Compustat identifier (anonymized)
seccik = SEC CIK Number (anonymized)
Casinoname = Name of the casino at which access takes place
NameBusinessEntity = Name of the overall business entity owning the casino
date = date of access (randomized)
time = time of access (randomized)
vhour = hour of access (randomized)
vminute = minute of access (randomized)

II. **Program Files**

Table_Code_FM.do: This is the STATA program that produces Tables 3, 4, and 5. The only dataset used in this code is Vegas_data_fm.csv.

Table_Code_FM.sas: This is the SAS program that produces the data backing Figures (1 – 4) and Tables 1, 2, A2, and A3. The only datasets used in this program are Vegas_data_fm.csv and Ipaccess_fm.csv.