SportsEconomics Perspectives, Issue 7

- What Drives Endorsement Earnings for Superstar Athletes?
- Is there a Consensus?: An Experimental Trial to Test the Sufficiency of Methodologies Used to Measure Economic Impact

SportsEconomics
The Science Behind the Business of Sports

www.sportseconomics.com

What Drives Endorsement Earnings for Superstar Athletes?

Athletes' endorsement earnings receive significant attention in the trade/popular press, but the academic literature on this issue is sparse. Thus, the purpose of the study was to examine factors that drive variation in endorsement earnings by superstar athletes.[1] Resulting analyses indicated the Positive Q Score measure (likability) was worth about $750,000-$1 million annually in earnings for each unit of the scale, while the Negative Q Score measure was statistically insignificant. One unit increases in Exposure and Familiarity were worth roughly $600,000 and $200,000, respectively. These findings can be used by athletes, agents, and sponsors to determine estimates, or fair-market value, for endorsement deals.

Introduction

The use of celebrities as endorsers continues to be a popular marketing strategy, as it is estimated that 15-25% of all advertisements in the United States feature celebrities. As such, approximately $1.6 billion is spent each year on athlete endorsements, with over 70% of that total going to only 100 athletes. Though research is somewhat mixed, it is generally accepted that endorsements, at worst, represent fair-value contracts, and many times can have positive pay-offs in brand-level sales and significant increases in the firm's stock returns.

Research Methods and Data Description

The two main sources of data used in this study are the top fifty salaries and endorsements figures of U.S. professional athletes as reported in *Sports Illustrated* for 2004-2012 (SI 50), and the Q Scores of the athletes on the SI 50 compiled by *The Q Score Company*. There are 155 unique athletes within the 450 observations from the *Sports Illustrated* database, and details of the data collection methods can be found with the SI Top 50. Q Score variables come from annual surveys with respondents drawn from a nationally representative, balanced panel of 2,000 respondents (Marketing Evaluations, Inc., n.d.). Each athlete is rated as either "One of my Favorites", "Very Good", "Good", "Fair", "Poor", or "Never Seen or Heard Before". One hundred and thirteen (113) of the 450 observations did not have a Q Score for that year, reducing total observations to 337. Table 1 includes a summary of salaries and endorsement earnings.

[1] This *SportsEconomics Perspectives* article is taken from the following article: Rascher, D., Eddy, T., & Hyun, G. (2017). What Drives Endorsement Earnings for Superstar Athletes? *Journal of Applied Sport Management, 9*(2).

Table 1

Endorsement Earnings by Sport

Sport	Count	Average Salary	Average Endorsement	Ratio of Endorsement to Total Earnings
Basketball	140	$16,198,507	$7,303,597	31%
Baseball	90	$18,329,972	$2,786,389	13%
Football	45	$16,677,110	$6,411,111	28%
Auto Racing	25	$7,388,727	$15,406,946	68%
Golf	21	$7,519,810	$50,314,286	87%
Boxing	8	$40,125,000	$3,468,750	8%
Tennis	6	$1,870,167	$23,500,000	93%
Cycling	2	$478,750	$17,000,000	97%

Variable Definitions

The dependent variable, Earnings, is simply each athlete's measure of endorsement earnings for a given year. The independent variables are defined as follows:

- **Salary** – The individual athlete's annual salary or earnings from their sport only.
- **Athletes** – This is the number of athletes for a given "team" playing at any one time (e.g., 5 for basketball, 9 for baseball, 1 for tennis).
- **Total Games** – The average annual number of contests played by the athlete. A boxer may be on television for 3 hours during a given year, while a tennis player may be on for 3 hours bi-weekly.
- **Exposure** – This is a separate measure of visibility created by dividing Total Games by Athletes. It is assumed that more games and fewer athletes in gameplay begets more camera coverage.
- **Total Familiar** – The Q Score Company's measure of the familiarity of the athlete. The average is about 64, meaning that 64% of the people in the survey recognized the athlete.
- **Positive Q Score** – This is the number of respondents who answered "One of my Favorites" divided by Total Familiar. A positive impact on endorsement earnings is expected.
- **Negative Q Score** – This is the number of respondents who answered "Fair" or "Poor" divided by Total Familiar. It is expected that this will have a negative impact on endorsement earnings.
- **Number of Years Professional** – This is the tenure of the athlete (averages 9.8 years). It may account for an athlete's ability to build up a following over time or establish a personality.
- **Sex** – This is an indicator variable with female = 1. Prior research has mixed findings on this so the a priori expectation is unknown.
- **Tennis, Baseball, Basketball**, etc. – these are indicator variables denoting each sport.
- **Year** indicator variables – These help control for time trends in the economy and endorsement market in particular.

Data Analysis

The research design echoes the value proposition of individual athlete endorsers, framed by explanatory factors including expertise, likability, familiarity, exposure, and demographic characteristics. The model examines revealed endorsement value, given that the data is only for actual endorsements that the athletes have contracted, and not all products they could endorse.

The initial analysis relies on ordinary least squares regression (OLS) and uses all of the available independent variables. Regression analysis helps separate out the incremental impact of each explanatory factor on annual endorsement earnings. Given that some of the athletes appear in multiple years of the data, cross-sectional time series analyses were conducted.

Results

The first model in Table 3 (Linear 1) uses fixed effects for each sport and does not include Athletes, Total Games, or Exposure. Eighty-six percent of the variation in endorsement earnings across the athletes is explained by this model. The key variables (Familiarity, Positive Q Score, and Negative Q Score) show that those athletes who are more familiar earn an additional ~$350,000 for each percentage point of increase in familiarity. Similarly, those with a positive Q Score earn over $500,000, all else equal, for each one percent increase in being selected "One of my Favorites." Interestingly, negative Q Scores did not have an impact on earnings. Linear 2 replaces the sport-specific indicators with structural characteristics of those sports: the number of athletes on the playing surface at a given time, the number of games/events during the season, and exposure (games divided by athletes). The exposure variable is highly significant, driving nearly $800,000 in additional endorsement earnings for each point increase. Positive Q Score is also highly significant ($900,000 increase for each additional point).

Linear 3 accounts for the significant correlation between Positive Q Score and Negative Q Score by using the difference between the two (Difference in Q Score equals Positive Q Score minus Negative Q Score). The correlation between the two variables can cause the coefficients on those variables to be misleading. Thus, creating this single variable removes any correlation (due to the existence of only one variable) and allows for a robustness check of the results. Estimating multiple models allows for a check on the robustness and stability of the results, and as shown, the effects of the key variables of interest are consistent across the models. Moreover, this new variable measures the gap between positive and negative responses about an athlete in the survey, allowing for a measure of relative likability to be tested. This new variable is highly statistically significant with a coefficient of about $500,000. The impact of the exposure variable is similar to that in Linear 2.

On average, an athlete appears in the data 3.3 times or years (minimum of one, maximum of nine). As shown in RE GLS 1 (random effects), most of the variation is explained by differences between athletes, not within the same athletes over time. The exposure variable is similar with a coefficient of about $600,000. Between 1 and Between 2 are between-group GLS analyses designed to control for Total Games and Athletes being correlated. However, the exposure variable is still worth around $500,000 for each unit increase in exposure.

Between 2 uses the Difference in Q Scores in place of Positive Q Score and Negative Q Score. The results are similar in that exposure and Difference in Q scores are associated with about $500,000 increases for one point changes in each variable. It is important to note that Sex contains only four observations that are not zero (male): Michelle Wie, Venus Williams, and Serena Williams (twice). Essentially, it is almost an indicator variable for tennis, but removing it from the analyses does not change the impacts of the other variables (other than for Tennis).

Table 3

Endorsement Regression Models

	Linear 1		Linear 2		Linear 3		RE GLS 1		Between 1		Between 2	
Number of Observations	337		337		337		337		337		337	
R squared	0.8612		0.6537		0.4977		0.6303		0.5646		0.4269	
Within							0.1557		0.0153		0.0254	
Between							0.6497		0.6227		0.4768	
F statistic	44.16		12.55		8.4				10.26		6.17	
Dependent Variable	Annual Earnings		Annual Earnings		Annual Earnings		Annual Earnings		Annual Earnings		Annual Earnings	
Salary	-0.0727664		-0.0541583		-0.1211005		-0.0675839		-0.1889063		-0.229571	
2004	-9368490	***	-8747977	***	-1958029		-5642857	***	-7371611	*	2708033	
2005	-5959112	***	-5608580	**	1238633		-2968983		-669589		9916280	*
2006	-5257905	***	-6378341	***	18369.9		-2659475	*	-3334994		2564015	
2007	-5663746	***	-7936107	***	-1977704		-2811779	**	-10000000	**	-849457	
2008	-3000212	*	-4650809	*	2040224		67331.83		-3182775		6204738	
2009	-2309285	*	-2771445		3036275		385675.4		-3361305		3717298	
2010	-3315467	***	-5370624	**	2090115		-933503.8		-2134021		10200000	*
2011	-467791.3		-1371024		4386305	*	1637541		281661.3		8281081	*
Sex	-27800000	***	-346587.4		4611028		-595588.9		630015.8		2840957	
Tennis	29900000	***										
Baseball	1049584											
Basketball	3567543											
Cycling	2445154											
Football	-7560961	***										
Golf	45200000	***										
Autoracing	4008837											
Positive Q Score	585705.1	***	901565.4	***			384826.4		1192314	***		
Negative Q Score	37536.91		160549.9				-148599.7		301113.3	**		
Difference in Q Score					488171.2	***					448196.5	***
Familiar	347044.6	***	117197.7	*			219599.4	***				
Number of Pro. Years	-880013.7	***	-475765.6	**	-157367.1		-276479.2		-663930	***	-332836.6	
Athletes			21584.19				-130270.6					
Total Games			-51377.01	***			-57451.32	***				
Exposure			791535.3	***	752972.9	***	603595	**	549886.1	***	539297.4	***
Constant	-15700000	***	-19100000	***	532545.4		-8491673		-18400000	***	2263081	

*Significance: * 10% level; ** 5% level; *** 1% level.*

Discussion

The purpose of this study was to investigate the impact of relevant predictors on superstar athletes' endorsement earnings. The model holds obvious practical implications, in that it could be used to determine a priori estimates (or future value) for endorsement deals, which could be beneficial for the companies that use athletes in advertising, as well as the athletes or sport agents who represent athletes' marketing interests. An athlete's relevant information can be inputted into the model, and the result can then be used to negotiate terms with interested brands. Of the variables in the model, Exposure, Positive Q Score (or Difference in Q Score), Familiarity, and sport played had significant effects on endorsement earnings. Specifically, if an athlete can improve their familiarity or likability by as little as 1 Q-score point, the model would suggest that they could see an aggregate increase of $750,000 - $1 million (or 1.5-2.0%).

Additionally, the findings in this study illuminate which characteristics brands believe are important in an endorser, and provide insight into what the market is willing to pay for those characteristics. Therefore, advertisers could use the models to identify prospective endorsers that may be undervalued by the market, but still have the desired personality attributes for a particular campaign. For example, an NFL athlete that scores highly on likability and familiarity may come at a lower price than a similar scoring athlete from golf or tennis, but could still bring as much value to the advertising campaign.

Additionally, it seems reasonable to suggest that the likability and familiarity of an athlete to the public may be within some degree of control for advertisers. Is it possible that Michael Jordan's fame grew to stratospheric heights, in part, because of the power of Nike's advertising campaigns and the Air Jordan brand, making him more familiar/likable to broader consumer segments? Would Peyton Manning score as highly on likability if not for his roles in humorous national campaigns that have become ingrained in popular culture? This proposition could be an important question for future research.

The increased goodness-of-fit when using the sport-specific variables shows that there are other factors that are not captured by exposure, number of athletes, or number of games/events per season (perhaps helmets covering the athletes' faces or the brutality of the sport, e.g., boxing and football). For example, the NFL receives arguably the most mainstream media coverage in the United States, yet NFL had a negative relationship with earnings. Future research should focus on gaining a better understanding of these between-sport differences.

Consistent with past research, athletes who played golf and tennis generally received higher endorsement earnings over the period in the study, but some of this positive relationship was driven by outliers (like Tiger Woods). It has been postulated that athletes in individual sports generate stronger sponsor impressions via television, and given that golf and tennis typically attract higher-income individuals, it is possible that companies are willing to pay more to endorse athletes in these two sports to gain access to these desirable target markets.

The models also highlight factors that are not predicting endorsement earnings among this sample of athletes. Although the number of years as a professional was the only direct measure of an athlete's quality (and was not statistically significant), the high explained variance suggested that athletic prowess is not playing a significant role in driving endorsement deals at the top levels of the market. Kobe Bryant could be an appropriate example here – he earns less endorsement money than his "peers", despite being arguably the top player of his era, because his

Q-scores are poor (likely stemming from his past legal issues). Thus, advertisers should be careful to not pay premiums for athletes solely because they demonstrate high levels of skill.

Additionally, the results of this study indicate that advertisers should consider long-term strategies that employ likable, up-and-coming athletes that have yet to attract national attention. These athletes will undoubtedly be cheaper than top stars, and their familiarity could be improved over time through their integration in a well-crafted advertising campaign. One recent campaign that may have exploited several of these possible inefficiencies was Verizon's ad series involving backup NFL quarterback Luke McCown. Exact figures were unavailable, but it seems likely that McCown was being paid significantly less than what a more well-known quarterback would demand. However, this clever campaign has received a significant amount of national attention, especially on social media. Thus, this endorsement appears to have been effective in communicating a mundane product feature (backup cellular towers), and it seems reasonable to suggest that this campaign has also improved McCown's familiarity.

It is important to note that there were no direct metrics included to measure fit between brand and athlete. Although the important fit-related concepts of expertise and athletic hero status would have been partially caught by the salary, familiarity, and positive Q-score variables, these three factors are clearly not equivalent. Given the high goodness-of-fit in the models presented here, perhaps fit is less important to companies when determining endorsement earnings/value than we might believe. Additionally, while the familiarity and the likability of the athlete are captured in the data, respondents' attachment to the athlete is not captured. To the extent attachment matters, the statistical analysis is not predicting a major effect, perhaps because it is already being caught in the Q Score information.

Finally, although more extensive than some of the studies in the literature, the current sample was limited to the individuals on *Sports Illustrated's Fortunate 50*, which only includes 102 athletes during the sample period (resulting in 337 athlete-years). Additionally, the endorsement earnings data from SI are estimates, and the method by which SI collects these data (relying heavily on sports marketing executives and agents) could lead to some inaccuracies. However, it seems likely that any issues with the estimates (likely overestimation) would be consistent across the sample; thus, while some of the quantified findings could be somewhat inflated, the relationships between the variables should not differ. Future research should also expand outside of this list with a more global sample that also considers retired athletes.

Author: Daniel A. Rascher, Ph.D. is President of SportsEconomics, LLC, and Professor and Director of the Sport Management Master's Program at the University of San Francisco. He has conducted over 150 sports business consulting engagements, and co-authored the textbook Financial Management in the Sport Industry.

Is There a Consensus?: An Experimental Trial to Test the Sufficiency of Methodologies Used to Measure Economic Impact

This research utilizes local Gross Domestic Product (GDP) of 383 Metropolitan Statistical Areas (MSAs) in the U.S. to determine whether historical methods in the academic literature to measure the economic impact of sports are sensitive enough to generate conclusive results.[2] An experiment is created and shows that commonly used methods fail to be able to detect the built-in-by-design injections of economic activity for the experimental group until very high levels of treatment of at least $300 million to $1 billion annually are present, thus providing evidence that Type I errors (rejecting a true null hypothesis) are likely to have occurred in some of the literature.

> *There are a lot of things economists disagree about, but the economic impact of sports stadiums isn't one of them. 'If you ever had a consensus in economics, this would be it,' says Michael Leeds, a sports economist at Temple University. 'There is no impact.' (Bergman, 2015, paras 1-2)*

As the above quote states, nearly all economists who have studied whether a facility, team, or major sporting event has a positive economic impact on a city or region have concluded that either there is no impact or that in a few cases there is a small positive impact (or a small negative impact). Economic impact is typically operationalized by measuring whether there is a net new change in the local economy (e.g., GDP, employment, number of firms, tax revenues collected, overnight stays at hotels) because of a new facility, team, or sporting event. The academic literature supporting this consensus historically utilized regression techniques on panel data consisting of cities (or MSAs) over a number of years with the use of indicator or dummy variables representing whether a market has a team, stadium, or hosted a major sporting event.

More recent research has narrowed the time periods (to quarters, months, or days [when using hotel occupancy or tourist arrivals]) and/or the geographic region (from MSAs and counties to census areas and zip codes) to enable more fine-grained analysis, as some authors have noted that the annual analysis on county-level data, for instance, could be too insensitive to detect what is attempting to be detected.

These regressions either use the level of GDP regressed on the dummy variables and control variables, or lagged dependent variables, or the growth in GDP as the dependent variable on the growth in the control variables along with the dummy variables of interest. The results often show the dummy variables as not being statistically significantly different from zero at conventional significance levels. Then, the authors tend to conclude that the existence of the team, stadium, or event has no net positive effect on the local economy.

However, industry practitioners often utilize intercept surveys at sporting events, as opposed to the *ex post* studies from the academic literature, to directly measure whether there is net new incremental money being spent in the municipality due to a new stadium, team, or event. If attendees to a sporting event come from out of town, spend

[2] This *SportsEconomics Perspectives* article is taken from the following article: Rascher, D., Hyun, G., & Nagel, M. (2020). Is there a Consensus?: An Experimental Trial to Test the Sufficiency of Methodologies Used to Measure Economic Impact *Journal of Applied Business and Economics, 22(12).*

money at local hotels and restaurants, and would not have spent that money in town had the event not taken place (i.e., they are in town because of the event and it does not take the place of another trip to that same town at some other point in time), then that expenditure is deemed to be an incremental gain to the local economy. This is generally the methodology used by industry practitioners. These practitioner reports often show quite large impacts, e.g., more than $100 million per year because of a new team or stadium, and tens of millions or hundreds of millions of dollars from major sporting events, like the Super Bowl. These can be ten times (or more) larger than what the academic sports economists claim is likely the true impact.

What explains the difference in the findings? The academic economists tend to claim that the industry reports either use the incorrect methodology and/or there are incentives to generate large impacts because the reports are often commissioned by teams, events, politicians and others who have a vested interest in showing large impacts.

While errors and embellishments in economic impact studies may certainly occur, an alternative to the results generated by academic economists is that their methods employed <u>may not be sensitive enough</u> to be able to measure the true underlying economic impact. Most of the early academic studies address the impact of teams in major sports leagues or major sporting events such as the Super Bowl or MLB All-Star Game on the host cities. In nearly all of those cases, the host cities are the largest such metropolitan areas in the United States, with annual GDP exceeding $250 billion. The average 2017 GDP of the top 50 largest MSAs in the U.S. exceeds $250 billion. Thus, if a team were to truly bring net new economic impact of $100 million annually to one of these markets, it would only be an increase of 1/2500 in GDP. This might not be detectable using annual panel data regressions. The average change in GDP from 2016 to 2017 in the top 50 U.S. markets (defined by size of GDP) exceeds $10 billion. Thus, even the change in GDP in a typical market is more than one hundred times the $100 million example, and thus might not be easily detected in panel regressions given the large variance in local GDP over time and across markets.

This article introduces a novel experiment in order to test whether the historical methodologies used in the academic literature are powerful enough or sensitive enough to measure true impacts of varying sizes. The experiment involves taking the existing data on GDP across markets and over time as given and then injecting various amounts of economic impact into randomly selected markets. Then, panel regressions are estimated to see if those economic impact injections can be detected. In other words, a randomly chosen ten percent of the markets will have millions of dollars added to their annual GDP for one year. The year is randomly chosen for each market (using a six-year panel). Thus, there is not a question about whether or not there is a true economic stimulus coming from hosting a major event because that stimulus has been forcefully added to that market. If the resulting panel regressions (with dependent variables using levels of GDP both with and without lagged dependent variables on the right-hand side of the equation, and also using growth in GDP) show no statistically significant impact, then the methods are not suitable for measuring these sizes of impacts. Variations on how many markets get the stimulus (10% up to 50%), the dollar size of the stimulus (ranging from $25 million to $1 billion per year), separate analyses for large, medium, and small markets, and stimuli designed to represent a team, are analyzed in order to be sure the findings are robust. The remainder of this article discusses the relevant literature, methods, data, and results.

Research Methods

Events

The methodology used in the present article is distinct from all of the previous literature concerning the economic impact of sports in that instead of testing whether an impact exists for facilities, teams, or actual sporting events, economic impact (or GDP of various amounts) is randomly added to actual GDP for a number of MSAs and then the typical regressions used in the literature are estimated to see if that impact is detectable. For instance, in one of the iterations approximately 10% of the markets are randomly chosen and $50 million is added to annual GDP in each of those markets. Treatments ranged from $25 million to $1 billion.

These models are tested for all 383 MSAs as a whole, and also separately for the top 50 markets by GDP (large markets), the next 100 markets (medium-sized markets), and the remaining 233 markets (small markets). Additionally, 30% and 50% of the markets are randomly chosen to host an event (not just 10% of the markets as described above). Moreover, larger treatments are administered: $100 million, $300 million, $500 million, and $1 billion. Finally, the data is re-created 20 times using different randomly selected markets and years. The results reported come from one of the sets of outcomes with any abnormalities described therein.

Teams or Stadia

The methodology above assessed markets that hosted a major sporting event either zero or one time during the six years of data. Much of the literature, however, is focused on sports facilities because often hundreds of millions of dollars of public money is used to pay for the construction of these facilities and industry stakeholders report large positive economic impacts that will ensue upon their completion. It is typically less controversial for an existing facility to host a sporting event (when presumably the costs to the public are much lower) than for a new venue to be built with public money. Of course, for cities or countries that host the Olympics or FIFA World Cup, respectively, significant construction projects are the norm. Yet, for an existing stadium to host an MLB All-Star Game, the public investment is much smaller. There are certainly costs to the public and local government when hosting a major event. These often include the effects of traffic congestion, overtime pay for police, fire, and other city services.

The only difference in analyzing events compared with facilities (or teams) is that new venues are built and then continue to be operational in the ensuing years. Thus, the indicator variable takes the value one for the opening year of a new facility and for all years beyond that, not just for one year as in the events analysis.

There are some more recent studies that have utilized quarterly or monthly data (often sales taxes collected as an indicator of spending), or daily hotel occupancy information, to measure economic impact. These shorter periodicities can be useful for investigating the impact of events, but less so for measuring the impact of facilities, which tend to be available for most of the year, or teams, which operate across multiple quarters and maintain yearlong business operations in the community. Thus, notwithstanding the same general principle that a given event may be too small relative to the size (and variance) of a local economy, the present analysis may leave open the possibility that the quarterly or monthly studies focused on short-term events are sufficiently sensitive enough to be able to detect the true underlying economic impact of most relevant events.

Data

The data come from the U.S. Bureau of Economic Analysis and provide annual GDP for each of 383 metropolitan areas within the U.S. from 2012-2017. Even the smallest market of the 383 measured by GDP in 2013, Grants Pass, Oregon, generates nearly $2 billion per year and has a population of about 37,000. Even if it hosted an event that brought in $2 million in new spending by visitors, that would still be about 1,000 times smaller than the local economy, thus making it hard to detect using a time series, panel data, or other statistical tests.

The largest market, New York-Newark-Jersey City, generated over $1.7 trillion in GDP in 2017, and that was $55 billion more than the year before. It is likely that any event or perhaps even any business (e.g., Amazon's HQ2) that operates in that area might not show up in a statistical analysis simply because the market is so large compared to any one event or business.

Results

<u>Events</u>

In general, the vast majority of the tests reject the null hypothesis that a true economic impact can be detected using the models and data that are commonly utilized in the literature. This is a classic Type I error in that a true underlying economic impact cannot be detected and thus is generally assumed to not exist even though it does exist, in this case by construction of the experiment.

Table 1 shows the results of applying a $50 million one-year treatment (a major event) to a random sample of 30% of MSAs, with the year chosen randomly within the six-year panel. As expected, the baseline models in the control groups (first, third, and fifth columns) show no significance from the EVENT variable. The control variables act as expected, with the effects of each year decreasing compared to the omitted year (2017). The effect of a $50 million event on the change in GDP shows a jump in the coefficient of about $60 million (consistent with expectations), but it is highly insignificant.

Table 1. Regressions Results for 30% of Markets Getting a $50 Million Event Treatment

	GDP Level	GDP Level with Treatment	GDP with Lagged	GDP Treatment with Lagged	Change in GDP	Change in GDP Treatment
Event Indicator	892	942	0.535	59.976	1.87	61.31
Lagged dependent variable	--	--	1.001 ***	1.001 ***	--	--
2012	-7983 ***	-7983 ***	--	--	--	--
2013	-6718 ***	-6718 ***	-537 ***	-536 ***	-546 ***	-545 ***
2014	-5010 ***	-5010 ***	-70	-69	-77	-76
2015	-3089 ***	-3089 ***	108	108	104	103
2016	-1794 ***	-1794 ***	-500 ***	-499 ***	-501 ***	-501 ***
Intercept	45763 ***	45763 ***	1739 ***	1735 ***	1803 ***	1800 ***
	---------- MSA fixed effects not shown ----------					
F	32.66	32.68	4311	4311	6.36	6.38
Number of observations	2298	2298	1915	1915	1915	1915

Statistical significance: *** - 1% level, ** - 5% level, * - 10% level. 2017 is the omitted year.

Not until the treatment reaches $500 million (with 30% of markets receiving it), does EVENT show up as being statistically significant, with a coefficient of $595 million (and p-value of 0.013) for the change in GDP model. It is not significant in the levels model. For the specification with 50% of markets receiving a treatment, statistical significance occurs at $300 million. Restricting the analysis to small markets (those outside of the top 150 in 2017

GDP), with 50% of those markets getting a $50 million treatment, the EVENT coefficient is statistically significantly different from zero and from the counterpart in the baseline regression.

To show the importance of the size of the markets and possible impacts, a $100 million Event in a large market (top 50) is not statistically significantly different from zero or the baseline coefficient. This is also true for a $300 and $500 million treatment. Not until the event generates $1 billion, does the model show statistical significance for top 50 markets.

Stadium or Team

Other than the controversy surrounding the Olympics, FIFA World Cup, and other major international mega-events, where significant public money is typically spent on construction projects related to the events, most of the criticism of sports industry economic impact analyses relates to public money being used for new stadia and arenas for a local team. This analysis also includes 432 separate regressions in order to cover a range of possible types of economic impacts.

Table 2 shows the results of providing a $500 million injection to local GDP for 30% of the 383 MSAs with the time period of the facility opening randomly assigned over the six-year period. As expected, STADIA is not statistically significantly different from zero in the baseline model because it is randomly chosen so should generally have no correlation to the dependent variable. Once $500 million is added to the treatment group of MSAs, the coefficient on STADIA rises by $500 million, as expected, but it is not statistically significant.

For the model utilizing the change in GDP, while the application of the treatment is seen in the coefficient rising from -$81 million for the baseline to $212 million for the experimental regression, both coefficients are not statistically significantly different from zero and from each other.

Table 2. Regressions Results for 30% of Markets Getting a $500 Million Facility Treatment

	GDP Level	GDP Level with Treatment	GDP with Lagged	GDP Treatment with Lagged	Change in GDP	Change in GDP Treatment
Stadium Indicator	-1035	-535	-80	213	-81	212
Lagged dependent variable	--	--	1.001 ***	1.002 ***	--	--
2012	-8251 ***	-8251 ***	--	--	--	--
2013	-6955 ***	-6955 ***	-554 ***	-502 ***	-563 ***	-513 ***
2014	-5169 ***	-5169 ***	-82	-37	-89	-46
2015	-3219 ***	-3219 ***	99	124	95	118
2016	-1866 ***	-1866 ***	-504 ***	-492 ***	-506 ***	-494 ***
Intercept	46,117 ***	46,117 ***	1764 ***	1694 ***	1827 ***	1772.128 ***
	---------- MSA fixed effects not shown ----------					
F	32.71	33.57	4311	4311	6.39	6.58
Number of observations	2298	2298	1915	1915	1915	1915

Statistical significance: *** - 1% level, ** - 5% level, * - 10% level. 2017 is the omitted year.

It is not until the treatment is $1 billion, meaning that the new stadium generates an annual economic impact of $1 billion, that the change in GDP specification (not the levels model) becomes statistically significant with a coefficient of $504 million and a p-value of 0.046. When 50% of the markets are in the experimental group, a $300 million treatment shows an economic impact of $373 million, significant at the 10% level. When restricting the analysis to medium-sized cities with 10% of the markets receiving a new stadium (or team) at some point during the six-year period, an annual economic impact of $100 million results in a z-score (comparing to the baseline coefficient) of 2.10, which passes the one-sided test.

Conclusion & Discussion

Over 800 different regression specifications were analyzed in order to determine if an economic impact injection into a local economy representing a major event, team, or facility was actually detectable using standard estimation methods. Models whose dependent variable was the change in local GDP from one year to the next unsurprisingly showed statistically significant results at lower levels of treatment. However, the thresholds (shown in Table 3) whereby a treatment becomes large enough to show statistical significance tend to be higher than what many of the industry reports claim is the economic impact of an event, team, or new facility.

Table 3. P-Values Showing How Large the Economic Impact of an Event Must be in Order to be Detected Using Historical Econometric Methods

10% of cities

	GDP	GDP with Lag	First Difference GDP
No Adjustment	0.814	0.435	0.437
Additional 50M	0.835	0.361	0.363
Additional 100M	0.855	0.296	0.297
Additional 300M	0.938	0.115	0.116
Additional 500M	0.978	0.035	0.036
Additional 1,000M	0.772	0.001	0.001

30% of cities

	GDP	GDP with Lag	First Difference GDP
No Adjustment	0.367	0.998	0.994
Additional 50M	0.341	0.802	0.798
Additional 100M	0.316	0.618	0.614
Additional 300M	0.228	0.136	0.135
Additional 500M	0.160	0.013	0.013
Additional 1,000M	0.056	**0.000**	**0.000**

50% of cities

	GDP	GDP with Lag	First Difference GDP
No Adjustment	0.630	0.436	0.434
Additional 50M	0.582	0.264	0.263
Additional 100M	0.536	0.146	0.145
Additional 300M	0.373	**0.005**	**0.005**
Additional 500M	0.245	**0.000**	**0.000**
Additional 1,000M	0.066	**0.000**	**0.000**

The most generous reading of Table 3 shows that if 50% of the markets host an event that is expected to generate $300 million to a market during a given year, then the methods used in the literature ought to be able to detect it. As an example, if a researcher were to analyze 50 cities over a 15-year period to see if hosting a Super Bowl (once per year) would drive a positive economic impact, the table shows that the event would have to actually generate around $500 million in order to result in a statistically significant coefficient.

There are vast differences between the largest and smallest MSAs in terms of GDP, population, etc. A more refined analysis that allows for separate models for each of the three sizes of MSA (i.e., Size 1 is top 50, Size 2 is next 100, and Size 3 is the remaining 233 markets) shows that in smaller markets the threshold size of an event for it to be detectable using the standard techniques in the literature is about $50 million. As shown in Table 4, for medium-sized markets, it is $500 million, and for large markets, it is $1 billion. Much of the previous, academic research has been focused on large markets hosting major events and has found that these events tend to have no discernible economic impact on those markets. Yet, according to the present study's results, those major events would have to generate $1 billion in net new spending in the community in order to generate positive and statistically significant results. Even the Super Bowl is generally not claimed by industry stakeholders to generate that size of an effect, meaning that many of the previous academic studies may have been insufficiently sensitive to be able to detect any actual economic impact (if there is one).

Table 4. P-Values Showing How Large the Economic Impact of an Event Must be in Order to be Detected, Disaggregated by Market Size

50% of cities by size

	GDP	GDP with Lag	First Difference GDP
No Adjustment			
Size 1	0.225	0.403	0.424
Size 2	0.082	0.342	0.641
Size 3	0.522	0.244	0.183
Additional 50M			
Size 1	0.220	0.377	0.397
Size 2	0.125	0.521	0.868
Size 3	0.057	**0.001**	**0.001**
Additional 100M			
Size 1	0.216	0.352	0.371
Size 2	0.185	0.739	0.893
Size 3	**0.002**	**0.000**	**0.000**
Additional 300M			
Size 1	0.200	0.264	0.279
Size 2	**0.618**	0.370	0.183
Size 3	**0.000**	**0.000**	**0.000**
Additional 500M			
Size 1	0.184	0.192	0.204
Size 2	**0.741**	0.035	0.012
Size 3	**0.000**	**0.000**	**0.000**
Additional 1,000M			
Size 1	0.149	0.077	0.083
Size 2	**0.017**	**0.000**	**0.000**
Size 3	**0.000**	**0.000**	**0.000**

The summarized results on the impact of a new stadium or team are shown in Table 5. The bar is even higher than for events, in that if 30% of the cities examined acquired a new team or facility, the actual impact would need to be around $1 billion annually in order for the methods traditionally used in the academic literature to detect it. If 50% of the cities examined opened a new facility, the actual impact would need to be around $300 million in order to justify the use of the standard techniques.

Table 5. P-Values Showing How Large the Economic Impact of a Facility Must be in Order to be Detected

10% of cities

	GDP	**GDP with Lag**	**First Difference GDP**
No Adjustment	0.492	0.800	0.803
Additional 50M	0.510	0.753	0.756
Additional 100M	0.529	0.707	0.710
Additional 300M	0.607	0.535	0.538
Additional 500M	0.690	0.388	0.389
Additional 1,000M	0.913	0.141	0.142

30% of cities

	GDP	**GDP with Lag**	**First Difference GDP**
No Adjustment	0.293	0.751	0.748
Additional 50M	0.317	0.841	0.838
Additional 100M	0.342	0.933	0.930
Additional 300M	0.455	0.704	0.707
Additional 500M	0.587	0.399	0.401
Additional 1,000M	0.972	**0.046**	**0.046**

50% of cities

	GDP	**GDP with Lag**	**First Difference GDP**
No Adjustment	0.746	0.373	0.373
Additional 50M	0.795	0.293	0.293
Additional 100M	0.844	0.226	0.226
Additional 300M	0.954	**0.064**	**0.064**
Additional 500M	0.755	**0.013**	**0.013**
Additional 1,000M	0.343	**0.000**	**0.000**

Hence, it is not surprising that an industry report might conclude that a new stadium in a top 50 market (which averages over $250 billion in annual GDP) would have an economic impact of $200 million per year, and the academic literature would conclude the opposite, that the same markets receives no net economic impact from a new facility based on panel regressions similar to those in this study because the thresholds to get a statistically significant result from those regressions are at $300 million or more. Accounting for the size of the markets (Table 6) shows that even if a new team or facility in a large market generates $1 billion per year in net new economic impact, the standard methodologies used in the academic literature will not detect it and will thus conclude that no impact exists. Small market economic impacts tend to become detectable at $50 million.

Table 6. P-Values Showing How Large the Economic Impact of a Facility Must be in Order to be Detected, Disaggregated by Market Size

50% of cities by size	**GDP**	**GDP with Lag**	**First Difference GDP**
No Adjustment			
Size 1	0.766	0.393	0.392
Size 2	0.478	0.712	0.707
Size 3	0.765	0.369	0.264
Additional 50M			
Size 1	0.774	0.381	0.379
Size 2	0.369	0.848	0.815
Size 3	0.369	**0.038**	**0.040**
Additional 100M			
Size 1	0.782	0.369	0.367
Size 2	0.276	0.988	0.928
Size 3	**0.036**	**0.001**	**0.003**
Additional 300M			
Size 1	0.814	0.323	0.322
Size 2	**0.065**	0.487	0.631
Size 3	**0.000**	**0.000**	**0.000**
Additional 500M			
Size 1	0.847	0.281	0.281
Size 2	**0.009**	0.160	0.294
Size 3	**0.000**	**0.000**	**0.000**
Additional 1,000M			
Size 1	0.929	0.194	0.194
Size 2	**0.000**	**0.002**	**0.015**
Size 3	**0.000**	**0.000**	**0.000**

The small number of times the academic literature has concluded that a sporting event or stadium (team) has generated a positive economic impact can be explained by the findings here and are not some unexpected abnormality, but instead driven by the context or size of the market in which teams participate, facilities are built, or events occur. Coates and Depken (2011), using monthly sales tax data, found that a season of Baylor University football games (held in Waco, Texas) generates an additional 1.8% to the local economy, or $175 million. Their finding is entirely consistent with the present results suggesting that smaller markets, like Waco, get discernible economic impacts from some sporting events or teams. Similar analyses of college football in smaller markets show positive and statistically significant results under some circumstances (Baade, Baumann, and Matheson, 2011).

The present research is also consistent with Agha's findings (2013). Her study focused on minor league baseball teams in 269 metro areas, some with populations lower than used in the present analysis. Noting that these smaller markets might allow for statistically significant findings unlike the majority of the previous academic research - which focused on the largest markets - Agha found that some minor league teams had positive and statistically significant effects of 0.2% to 0.7% of local per capita income, or $67 to $118 in per capita income. This translates to about $30 million on average for each of the 233 small markets in the present data, which is consistent with the findings presented in Table 6. In other words, minor league teams in small markets that generate at least $30 million per year in economic impact may marginally be detected as being statistically significant. In addition, Agha found significance for some levels of minor league baseball in many markets, but not across the board. In other words, there may be positive impacts from more of the markets studied by Agha, but those impacts could be $20 million or less and thus not detectable by the standard academic methods utilized.

Given the large economies of the cities that tend to host major teams or sporting events, this paper demonstrates that the likelihood of committing a Type I error, not finding a positive economic impact when in fact there is one, is substantial. This can also be seen by noting the often large size of the confidence intervals around the coefficients of interest. E.g., in Baade, Baumann, and Matheson (2008) on the impact of major sporting events, Table 4 shows a coefficient on hosting the Super Bowl (with the dependent variable as the natural log of local taxable sales) to be 0.0336, a standard error of 0.0312, and a t-statistic of 1.08. The upper bound of a 95% confidence interval (not shown) would be 0.0948, implying that it also cannot be rejected that hosting the Super Bowl increases taxable sales by nearly 10 percent. In other words, the precision is very low for the sort of conclusions drawn.

More detailed and fine grained methods, such as the use of monthly data (Coates and Depken, 2011; Baumann and Matheson, 2018), daily hotel occupancy rates or tourist arrivals (Baumann and Matheson, 2017; Chikish, Humphreys, Liu, and Nowak, 2019) or a focus on smaller geographical regions is warranted (Feng and Humphreys, 2012).

These types of studies are occurring with more frequency, but will still have similar issues to solve albeit smaller. In other words, if monthly data is likely to provide a context that is twelve times smaller, there is still the possibility that an event impact will be subsumed by the size of the local monthly economy. Thus, future research should create similar guidelines for monthly data or daily occupancy rates. Additionally, shorter time periods have other potential problems such as if the timing of when monthly taxes are paid occurs months after an event (Coates and Depken, 2011), or if payments by event owners to local vendors occurs 90 days after an event (a personal communication with a major event owner by one of the authors confirms that this is often the case). The use of daily occupancy rates is even more fine-grained, but also may suffer from other measurement issues such as whether visitors stay in facilities captured by the data (e.g., is Airbnb data included, or do many visitors stay at a friend's residence?), or whether the location studied (e.g., within 4 miles of an arena) is where people stay when coming from out of town to attend an event at the arena. Yet, these same sets of data with specified periodicities and geographies can also be utilized to create threshold impacts that the models can sufficiently detect. Moreover, industry reports that attempt to account for the various errors presented by Crompton (1995) should not so easily be dismissed as clearly being incorrect simply because the results differ from the previous findings in the academic literature.

The present study demonstrates that much of the early seminal academic literature uses statistical analyses that are not sensitive enough to be able to detect economic impacts in the hundreds of millions (or even billions) of dollars in the large markets that tend to host major sports teams and events. This research helps provide guidelines for the size of events/teams/facilities that are detectable using standard techniques, and provides one explanation for the gap between the academic literature and industry reports that generate opposite findings. An industry report showing $200 million in annual economic impact from a stadium might be a fair measure and yet that amount would not be detectable in the academic literature. Further research should focus on these same types of experimental trials focused on monthly data and daily occupancy rates. Moreover, testing any econometric models on any topics, not just sports economics, that fail to find statistically significant coefficients using experimental trials of this sort (where impacts are randomly injected into the dependent variable) can help provide thresholds and guidelines of how Type I errors may occur.

Author: Daniel A. Rascher, Ph.D. is President of SportsEconomics, LLC, and Professor and Director of the Sport Management Master's Program at the University of San Francisco. He has conducted over 150 sports business consulting engagements, and co-authored the textbook Financial Management in the Sport Industry.