

The Future of Nature

Acadia National Park and visitation predictions

Oliver Bruce

June, 2008

College of the Atlantic

Class: Introduction to Econometrics with Davis Taylor

Introduction

Acadia National Park (ANP) was created in 1916. Located on Mount Desert Island in Maine, the park protected Cadillac Mountain, the highest point on the Eastern Seaboard, as well as large sections of woods that gave rise to hiking and carriage trails in order to preserve them for future generations. Now encompassing over 45,000 acres, the park has grown to include various off island sections, and is the 10th most popular National Park in the United States, attracting over 2 million visitors a year.

Acadia however, like most other National Parks, face declining numbers of visitors following its peak in visitation in 1997. A talk with a park ranger will yield complaints about increasing numbers of people resorting to electronic means to entertain themselves, changing holidaying trends and disconnection with nature as being at the root of this decline, yet no conclusive study has been published. While many of the variables which are speculated to influence visitor numbers are difficult to measure, there is sufficient evidence to analyze the historical quantitative factors that have driven visitor numbers in the past.

This study examines the variables that are theorized to impact park visitor numbers, and proposes an equation that can be used to forecast park visitation numbers in the years to come. The intention is that this can be used to aid in policy setting and long term strategic planning in ANP so that the parks can best serve the needs of their visitors in the years to come.

Section 1: Theory

I've been unable to find definitive literature on variables that affect visitation in the national parks and as a result, I've utilized a number of different sources when formulating theory. Primarily, I've used an interview with Charlie Jacobi, ANP's Resource/Visitor Use Specialist, early research results from the University of Arizona on visitation to the National Parks in their region and a collection of news articles. From these sources, I've identified 6 factors that theoretically influence visitor numbers. They are:

Factor #1 - Increasing gas prices decrease number of visitors to parks

With changes in the prices of gasoline at the pump, tourists are likely to adjust their traveling patterns - driving farther when the cost of travel is low and shorter distances when prices are high. According to graduate student Srinivasa Ponnaluru and Professor George Frisvold of University of Arizona, "annual visits fall by 2.3 percent for every 10 percent increase in the price of gasoline," for National Park visitation in the Southwest of the United States (see Bibliography). These trends appear to be echoed by several groups throughout the USA, not least among them, the Travel Industry Association (TIA). The Association speculates that for the 2008 summer, due to increased fuel costs travelers will decrease longer distance vacation travel plans and to holiday closer to home. While this may appear good for increasing ANP numbers, because of its relative proximity to the large population bases of New England, the distances may still be prohibitive. Bangor Daily News (see Bibliography) reported that for this year (2008) we will see a decrease in the amount of traffic heading towards ANP due to the high price of gasoline.

Factor #2 - Weather patterns

Charlie Jacobi pointed out that Acadia is a “short visit” park, and that the majority of those who visit will only come for a few days. Charlie suggested that because of this visitor dynamic, weather was a high impact variable on the decision of closer domestic visitors who could easily reschedule their trips in the event of a bad weather forecast. This variable could be measured both with the favorable weather conditions, like average high temperatures and with less desirable ones, such as rainfall.

Factor #3 - Highway miles increase access to national parks

With the spread of the US interstate system, getting to ANP became a significantly easier task. It significantly reduced travel time by road overall, and gave citizens of nearby states to the opportunity to visit the park for shorter visits, which hadn't been possible before. While there was no published literature located on the topic of the interstate increasing the prospects of the tourism industry, this was another point that Charlie Jacobi suggested historically might have caused a positive impact in increasing access to Acadia National park.

Factor #4 - Population

This factor was taken from the preliminary research results out of University of Arizona that appeared in the CLIMAS Newsletter (Volume 8, Number 2, Nov. 2005). The research was investigating the economic and environmental factors that had appeared to affect visitation in nearby parks. The size of the population base from which to ‘catch’

visitors is seen to be correlated to the visitor statistics of the park. Any change in the surrounding population will probably result in an effect on the numbers of visitors.

Factor #5 - Park entry costs

This factor was suggested during the conversation with Charlie Jacobi as being a potential factor influencing who comes to the park, especially for the population who live within an hour's drive to Acadia. The article, "Effects of per-vehicle entrance fees on U.S. National Park visitation rates" by Seth Factor supports this view. For those visitors that travel from further away, the cost of entry appears nominal in comparison to their other travel expenses. Charlie made the point however, that the cost of entry has gone up most years since 1996, and that it has had the effect of alienating some park regulars who lived nearby.

Factor #6 - Gross Domestic Product

The final factor also came from the preliminary research results from University of Arizona, who found that economic activity may influence park visitation. The reasoning behind this is that having a strong economy drives up incomes, which in turn, drives up money spent in tourism. Measuring economic growth, we can expect there to be a correlation with visitor park statistics over time.

Conclusion

These six factors are by no means the comprehensive list of variables to cover all possible factors influencing the decisions of potential visitors to ANP. It is impossible to

accommodate for every potential reason as the majority of them will be statistically insignificant. Thus, the above represent the 'heavy hitters'; those variables that are suggested to have widespread impacts.

Section 2: Specification and Hypothesizing Expected Signs

Firstly, for this particular model, we shall attempt to use a time series regression to calculate our equation. Time series are the most accurate method to look at relationships between variables such as annual visitation data. Secondly, it is important to select variables that are both reflective of the issue we are attempting to measure, and that has sound reasoning for their inclusion. For this reason, I have listed below variables that could potentially influence our dependent variable and what we would expect as their signs.

Variable	Abbreviation	Expected sign	Measurement unit
Park Visitation	RECVIS		Visitors per season
GDP	GDP(2000)	+	US\$ (2000)
Precipitation (July)	PRECIP	-	Inches of rain
Temperature (July)	TEMPAV	+	Degrees Fahrenheit
Interstate system	HIGHWAY	+	Cumulative miles
Population	POP	+	People
Gas Prices	GAS	-	US\$ (2006)
Park Entry costs	ENTRY	+/-	US\$ (2006)
Dummy for Visitors	dummyvisit	-	post 1990 = 1

LEFT HAND SIDE VARIABLE

Park Visitation (RECVIS)

My dependent variable is number of visitors to ANP in a given year. The park is technically open all year round, while the high season is defined as being between April 15 to December 1, when all roads and visitor centres in the park are open.

RIGHT HAND SIDE VARIABLES

Gas Prices

This variable tracks the pump cost of gasoline in the USA over a historical period. I've included it because its a measure of the cost of travel, which influences decisions to come to ANP. We can reasonably expect the sign to be negatively correlated, in line with the theory that with an increase in fuel costs, there is decreased motivation to travel long distances. While this could be countered by saying that high cost of travel means people look for destinations closer to home where they can vacation, and would pick ANP because of its relative proximity, I still expect an overall decline in the number of visitors.

Population

Population was included as a variable because general population for the area around National Parks makeup the main portion of the visitors. The number of people with easy access to the park can be expected to be positively correlated with the visitor figures. While park access is more dependent on various other variables as well (access to interstate highways, gas prices and disposable income), the size of the population could also have an impact. This variable had weaker specification reasoning and as a result was not as strong as others. It was included in the first regressions however, in case there was change statistical significance.

GDP

This variable measures the Gross Domestic Product annually, and divides it by the population in order to get GDP per capita. This gives some indication of (though, is not) average incomes. This can be expected to be positively correlated because with an increase in incomes, there can be an expected rise in discretionary spending such as tourism. I've pegged the GDP income to adjust for inflation over time so that the variable follows a variable trend. I have also lagged the GDP data a year, to reflect the fact that often slowing growth in a GDP isn't reflected in discretionary spending until the following year.

Precipitation

This variable monitors the July average for precipitation in inches. Given that the key features of ANP are located outside, Charlie Jacobi suggested that weather may play a part in peoples decisions to come to the Park. We can expect that the coefficient will be negative given the dampener rain puts on trips outdoors (forgive the pun). While we could have chosen to measure the cumulative rainfall throughout the season, I knew there was going to be challenges with obtaining data, and so I limited it to the most popular month on average for the year.

Temperatures

This variable accounts for the opposite end of the scale when looking at weather by measuring the average temperatures for the month of July. We can expect this coefficient to be ambiguous, as temperatures that are low on average could have a negative impact on the number of visitors, while heatwaves could also be expected to

have the same effect. Upon the suggestion of Charlie Jacobi however, we will include it in regressions to test for statistical significance.

Park Entry Price Costs

This variable is the inflation adjusted cost of entry that visitors need to pay to get into the park. While it would be reasonable to expect that the cost of entry would be negatively correlated - that with an increase in cost, came a decrease in demand for visitation - the cost is nominal (US\$20) for people who are traveling from further away when other travel costs are taken into consideration. Thus, we can expect this variable to be ambiguous.

Highway Miles completed

This variable measures the amount of interstate highway that was paved historically. As the primary mode of transportation is the automobile and the role of the interstate system is to facilitate long distance movement of automobiles, measuring the interstate gives us a reasonable indication of ease of movement for the US population. We can expect that with an increase in the number of miles of interstate, there will be a positive correlation in the numbers of visitors that choose to visit ANP.

Section 3: Data

It was a challenge to locate the data for the variables above, and in many cases, I ended up having to compromise on quality or accuracy. In all cases however, I am satisfied that the datasets that I have found adequately represent the variables that I am hoping to measure.

RIGHT HAND SIDE VARIABLES

Obtained from	Observations	Multicollinearity
www.nature.nps.gov/stats	1919-2008 = 89 obs	Unsure

This variable measures the number of people supposedly visiting ANP over a given season. I got my figures from the National Park Service online database on the suggestion of Charlie Jacobi (Interview, April 30th, 2008). The data has two primary weaknesses, firstly, that the method used to measure park visitation is rather unreliable, and secondly, has had several large changes to formula over the sampling period reducing consistency of the information. Generating park visitation statistics is done by first counting the number of cars that pass by a particular point on the loop road that runs around the park. That figure is then put into a formula that extrapolates on the number of cars that were in the park and didn't pass the counting device, and the number of occupants in all cars. The formula used is calculated from survey results taken at various points around the park during the summer on travel patterns. This results in only having purely estimated visitor numbers with no ability to provide verification.

Secondly, this formula they use however, has not been consistently the same over time, and has produced some questionable results in the past. Charlie Jacobi pointed out the change in the formulas used to calculate visitations between 1989 and 1990 resulted in a 'drop' in estimated numbers from 5.4 million to only 2.3 million. Its not that there really were differing numbers in actual visits, simply that the measured number showed large variation. The second weakness has motivated me to include a dummy variable (named dummyvisit) in the equation for pre- and post-1990 years to account for the differences in the dependent variable.

LEFT HAND SIDE VARIABLES**- Gas Prices**

Obtained from	Observations	Multicollinearity
http://www.swivel.com/data_columns/show/2690243	1970-2006 = 37	GDP

Historical gas prices were tied to 2006 prices to keep in line with the other variables which had been inflation adjusted. This source was chosen because it was the sample size with the most years available. I would have liked data for even further back, but the earliest I could find was from 1970. This dataset was the limiting variable in the equation, and the reason that we only have 37 full observations. The data set is from Swivel.com, a data-sharing website, and the source was not listed, and thus I couldn't certify it. I countered this, however, by crosschecking it with data from smaller, verified samplings and found it to be suitably accurate.

I expected a correlation with GDP, after lagging the figures a year behind to take into account the cost of fuel on productivity. I didn't expect any further collinearity with other variables.

- US Population

Obtained from	Observations	Multicollinearity
Historical Statistics for the World Economy: 1-2003 AD by Maddison	1951-2007 = 57 obs	GDP (2000), HIGHWAY, ENTRY

This measures the overall population in the United States over time. The data comes from a database of worldwide historical GDP and population statistics from Angus Maddison, Emeritus Professor at the University of Groningen in the Netherlands (see Bibliography). The database is considered to be one of the most comprehensive, publicly available records at the present time. The data is taken from textbooks and government records to track development for all countries in these

metrics since the year 0. While this may not have been the most appropriate of variables given the theory that the park pulls from its nearby population, historical data of New England populations was nigh on impossible to come by.

I expected multicollinearity with GDP, HIGHWAY and ENTRY because they were all expected to grow in line with POP.

- Gross Domestic Product

Obtained from	Observations	Multicollinearity
Historical Statistics for the World Economy: 1-2003 AD by Maddison	1951-2007 = 57 obs	HIGHWAY, POP

This variable measures economic development over time in the United States.

Like US population statistics above, its from the database of historical GDP and population statistics from Angus Maddison. The figures are adjusted for inflation to year 2000 US dollars to account for inflation. While this isn't necessarily the best metric to measure the performance of an economy, or the level of disposable income an individual may have, its the most precise tool we have and we'll have to make do with such a metric.

I expected imperfect multicollinearity with GDP and HIGHWAY - two variables that would grow exponentially.

- Precipitation

Obtained from	Observations	Multicollinearity
<u>National Climatic Data Center</u>	1970-94, 1999-2007	None

This variable measures the amount of rainfall during the month of July, one of the most popular months of visitation at Acadia. Unfortunately, we were unable to get sufficient historical data from the weather stations in ANP, Bar Harbor, or even Ellsworth, and so the data is coming from the weather station at Bangor International

Airport (BIA). BIA is located close enough that the variation in the weather between there and ANP is small enough to have this data remain relevant.

Unfortunately, this dataset is also incomplete, with a gap in measurement between 1995 and 1998 that cannot be filled. This is particularly problematic as there cannot be gaps in time series regressions - my intended model of regression. As a result, I had to shift to using a cross-sectional regression model, which drives up the correlation, though not significantly. It is adequate to predict this model, but we need to be conscious of the fact that the model isn't perfectly suited to the data type.

I did not expect any form of multicollinearity for this variable.

- Temperatures from July

Obtained from	Observations	Multicollinearity
National Climatic Data Center	1970-94, 1999-2007	None

This variable also tracks weather statistics, but looks at the more positive side of weather - the average temperature for July. I ran into the same problems with this dataset as I did with Precipitation above. Because of the gap in the data, we were restricted from running a time series regression. As with the the variable above, I would have preferred to get relevant Acadia statistics, but as this was all that was available, we had to make do.

I did not expect any form of multicollinearity for this variable.

- Park Entry Cost

Obtained from	Observations	Multicollinearity
http://hdl.handle.net/10161/397	1968-2007 = 40 obs	None

I obtained this data from the article "Effects of per-vehicle entrance fees on U.S. National Park visitation rates" by Seth Factor cited above. This data set is not full, only

going back as far as 1996, when Congress changed the fee structure for National Parks significantly. Before this time, Charlie mentioned the park charged no effective fees, and maintained a far more relaxed approach to non-paying users. The fees in this dataset have been pegged to 2006 dollars, and thus are adjusted for inflation.

I did not expect any form of multicollinearity for this variable.

- Highway miles

Obtained from	Observations	Multicollinearity
http://www.publicpurpose.com/hwy-intnmiles.htm	1956-1995 = 39 obs	POP,

This dataset from publicpurpose.com attributed their data to the Federal Highway Administration, though I was unable to verify their figures. I would have preferred to have had the amount of paved I-95, the highway that connects Bangor, and by extension Acadia, to the rest of New England. Unfortunately, that information wasn't available and so I had to settle for overall miles constructed. These figures in many ways are better, because the value of the interstate system is in the network, not in the road. The network connected a huge number more people to the I-95, making access to ANP all the more easier.

I expected imperfect multicollinearity with HIGHWAY from POP and GDP, as two other variables that grew exponentially and were two variables that were tangentially connected to the development of the Interstate System.

Figures I wanted, but couldn't get

There were two major variables that I was hoping to obtain on top of this, but was unable to find:

Obesity

Charlie Jacobi cited rates of obesity as being a potential variable that could be influencing the numbers of visitors to Acadia National Park. I would have like to follow this up, however, I was unable to locate sufficient historical data for this.

Changing means of entertainment.

Charlie also cited the dramatic increase in electronic media as a means of entertainment as being one of the key drivers in driving down park visitation statistics. Despite making several attempts (such as measuring the size of the video game industry, number of consoles, size of film sales for that year etc.) this was nearly impossible to quantify, and thus I chose to exclude it from the equation.

Section 4: Estimations

Because of the gaps in data with my weather data, I used a cross sectional model to run my regressions. I ran my initial regression without the dummy variable to see whether there was any statistical significance. The results were as follows:

WITHOUT DUMMY VARIABLE

Dependent variable:

Y = Recreational Visitors (NPS figures)

Characteristics:

Recreational Visitors (NPS figures)

First available observation = 1

Last available observation = 57

First chosen observation = 20

Last chosen observation = 56

Number of usable chosen observations: 37

Subsample characteristics:

Minimum value: 2.0514840E+006

Maximum value: 5.4409520E+006

Sample mean: 2.8740236E+006

This variable is integer valued.

A discrete dependent variable model might be more suitable!

X variables:

X(1) = GAS

X(2) = GDP 2000

X(3) = POP

X(4) = PRECIP

X(5) = ENTRY

X(6) = TEMPAV

X(7) = HIGHWAY

X(8) = GDP [Lag1]

X(9) = GDP [Lag2]

X(10) = 1

WARNING: The effective degrees of freedom is only 27.
Therefore, the estimation results may be unreliable!

Model:

$Y = b(1)X(1) + \dots + b(10)X(10) + U$,

where U is the error term, satisfying

$E[U|X(1), \dots, X(10)] = 0$.

OLS estimation results

Parameters	Estimate (S.E.) [p-value]	t-value (H.C. S.E.) [H.C. p-value]	H.C. t-value
b(1)	-705318.2691056 (431544.18024) [0.10217]	-1.634 (412475.49964) [0.08727]	-1.710
b(2)	467.8411912 (1347.68794) [0.72848]	0.347 (1410.09103) [0.74005]	0.332
b(3)	-0.0078486 (0.10064) [0.93784]	-0.078 (0.12514) [0.94999]	-0.063
b(4)	58606.7984409 (90287.24495) [0.51626]	0.649 (86791.58874) [0.49951]	0.675
b(5)	-60793.8030088 (54129.94092) [0.26139]	-1.123 (43530.22455) [0.16254]	-1.397
b(6)	25357.4726693 (70389.27488)	0.360 (58442.52918)	0.434

		[0.71866]	[0.66437]	
b(7)	204.4278745	1.700	1.313	
	(120.23260)	(155.64658)		
	[0.08908]	[0.18904]		
b(8)	-2.0156078	-0.001	-0.001	
	(1846.46585)	(1638.02185)		
	[0.99913]	[0.99902]		
b(9)	-604.7843389	-0.434	-0.522	
	(1394.79877)	(1159.18368)		
	[0.66458]	[0.60186]		
b(10)	-2387949.7921765	-0.188	-0.148	
	(12686163.68594)	(16081704.71099)		
	[0.85069]	[0.88196]		

Effective sample size (n): 33
 Variance of the residuals: 501083803796.409
 Standard error of the residuals (SER): 707872.7313553
 Residual sum of squares (RSS): 11524927487317.4
 (Also called SSR = Sum of Squared Residuals)
 Total sum of squares (TSS): 21135816246448.2
 R-square: 0.4547
 Adjusted R-square: 0.2414

As you can see, the Adjusted R-square is not particularly healthy, and the majority of the t-stats are unacceptably low - in all, an inappropriate estimated regression. For that reason, I ran the regression with the dummy variable included to see what difference it made. The results are as follows:

WITH DUMMY VARIABLE

Dependent variable:
 Y = RECVIS

Characteristics:
 RECVIS

First available observation = 1
 Last available observation = 57
 First chosen observation = 20
 Last chosen observation = 56
 Number of usable chosen observations: 37

Subsample characteristics:
 Minimum value: 2.0514840E+006
 Maximum value: 5.4409520E+006

Sample mean: 2.8740236E+006

This variable is integer valued.

A discrete dependent variable model might be more suitable!

X variables:

X(1) = GAS

X(2) = POP

X(3) = ENTRY

X(4) = PRECIP

X(5) = TEMPAV

X(6) = HIGHWAY

X(7) = GDP [Lag1]

X(8) = dummyvisit

X(9) = 1

WARNING: The effective degrees of freedom is only 28.

Therefore, the estimation results may be unreliable!

Model:

$Y = b(1)X(1) + \dots + b(9)X(9) + U$,

where U is the error term, satisfying

$E[U|X(1), \dots, X(9)] = 0$.

OLS estimation results

Parameters	Estimate (S.E.) [p-value]	t-value (H.C. S.E.) [H.C. p-value]	H.C. t-value
b(1)	-561638.7720982 (223504.51674) [0.01198]	-2.513 (181644.59946) [0.00199]	-3.092
b(2)	0.0546882 (0.03633) [0.13223]	1.505 (0.03711) [0.14062]	1.473
b(3)	-104313.4049254 (34298.33523) [0.00236]	-3.041 (41893.63437) [0.01278]	-2.490
b(4)	7143.3309142 (42998.83257) [0.86806]	0.166 (49834.50989) [0.88602]	0.143
b(5)	28809.5040959 (33952.58135) [0.39615]	0.849 (21770.66737) [0.18573]	1.323
b(6)	15.7154958 (65.02449) [0.80902]	0.242 (89.02180) [0.85987]	0.177
b(7)	18.4091936 (428.00735)	0.043 (473.99981)	0.039

		[0.96569]	[0.96902]	
b(8)	-2660943.4607906	-9.052	-7.133	
	(293962.82781)	(373030.65573)		
		[0.00000]	[0.00000]	
b(9)	-10885756.6444191	-2.119	-2.144	
	(5137169.46856)	(5077001.95462)		
		[0.03409]	[0.03202]	

Effective sample size (n): 33
 Variance of the residuals: 116998243811.167
 Standard error of the residuals (SER): 342050.06038761
 Residual sum of squares (RSS): 2807957851468
 (Also called SSR = Sum of Squared Residuals)
 Total sum of squares (TSS): 21135816246448.2
 R-square: 0.8671
 Adjusted R-square: 0.8229

As we can see, adding the dummy variable to adjust for the change in the measurement system significantly increased the statistical significance of the regression. However, while the Adjusted R-square is much higher (0.8229), we still have a large number of insignificant statistical results with low t-stats and a low number of degrees of freedom (28). PRECIP, TEMPAV and GDP (2000) were showing up as insignificant. Simply as an experiment, I multiplied the weather variables together, and logged GDP growth to see if I could increase their significance. The results were as follows:

REGRESSION WITH TEMPAV X PRECIP

Dependent variable:

Y = RECVIS

Characteristics:

RECVIS

First available observation = 1

Last available observation = 57

First chosen observation = 20

Last chosen observation = 56

Number of usable chosen observations: 37

Subsample characteristics:

Minimum value: 2.0514840E+006

Maximum value: 5.4409520E+006

Sample mean: 2.8740236E+006

This variable is integer valued.

A discrete dependent variable model might be more suitable!

X variables:

X(1) = GAS

X(2) = POP

X(3) = GDP [Lag1]

X(4) = LN[GDP(2000)]

X(5) = PRECIP x TEMPAV

X(6) = dummyvisit

X(7) = ENTRY

X(8) = HIGHWAY

X(9) = 1

WARNING: The effective degrees of freedom is only 28.

Therefore, the estimation results may be unreliable!

Model:

$Y = b(1)X(1) + \dots + b(9)X(9) + U$,

where U is the error term, satisfying

$E[U|X(1), \dots, X(9)] = 0$.

OLS estimation results

Parameters	Estimate (S.E.) [p-value]	t-value (H.C. S.E.) [H.C. p-value]	H.C. t-value
b(1)	-395663.1504519 (292332.13171) [0.17591]	-1.353 (237072.31848) [0.09513]	-1.669
b(2)	0.0474934 (0.03878) [0.22068]	1.225 (0.04679) [0.31010]	1.015
b(3)	-237.5855946 (547.47389) [0.66431]	-0.434 (428.08231) [0.57889]	-0.555
b(4)	2784606.3121753 (3887773.90760) [0.47384]	0.716 (3688357.70977) [0.45027]	0.755
b(5)	131.3532130 (627.66736) [0.83424]	0.209 (793.95242) [0.86860]	0.165
b(6)	-2631361.8748222	-8.794	-7.016

		(299227.81339)	(375048.52701)
		[0.00000]	[0.00000]
b(7)	-109079.9453279	-3.217	-2.706
	(33904.95977)	(40310.36835)	
	[0.00129]	[0.00681]	
b(8)	-42.6229227	-0.431	-0.389
	(98.93745)	(109.58940)	
	[0.66661]	[0.69733]	
b(9)	-27800567.0566690	-1.058	-1.272
	(26264440.31497)	(21847739.05725)	
	[0.28983]	[0.20321]	

Effective sample size (n): 33
 Variance of the residuals: 117760657073.907
 Standard error of the residuals (SER): 343162.72681325
 Residual sum of squares (RSS): 2826255769773.77
 (Also called SSR = Sum of Squared Residuals)
 Total sum of squares (TSS): 21135816246448.2
 R-square: 0.8663
 Adjusted R-square: 0.8217

Again, as you can see, we have a high Adjusted R-square, yet largely insignificant t-stats. Rather than continue experimenting to see if we could find something, I decided to drop those variables (PRECIP, TEMPAV and POP) as the theoretical reasoning behind their inclusion wasn't particularly strong. While GDP [Lag 1] had a very low t-stat, its theoretical reasoning was also stronger, and was worth keeping in the regression at this stage. The case was similar with HIGHWAY. That left us with 4 variables that were either statistically strong, or possessed sound reasoning for remaining in the regression: GDP [LAG1], GAS, ENTRY and HIGHWAY. I ran another regression:

STRIPPED DOWN -KEY VARIABLE- REGRESSION
 Dependent variable:
 Y = RECVIS

Characteristics:

RECVIS

First available observation = 1

Last available observation = 57

First chosen observation = 20

Last chosen observation = 56

Number of usable chosen observations: 37

Subsample characteristics:

Minimum value: 2.0514840E+006

Maximum value: 5.4409520E+006

Sample mean: 2.8740236E+006

This variable is integer valued.

A discrete dependent variable model might be more suitable!

X variables:

X(1) = GAS

X(2) = ENTRY

X(3) = HIGHWAY

X(4) = GDP [Lag1]

X(5) = dummyvisit

X(6) = 1

Model:

 $Y = b(1)X(1) + \dots + b(6)X(6) + U$,where U is the error term, satisfying $E[U|X(1), \dots, X(6)] = 0$.

OLS estimation results

Parameters	Estimate (S.E.) [p-value]	t-value (H.C. S.E.) [H.C. p-value]	H.C. t-value
b(1)	-640104.2469797 (183595.21257) [0.00049]	-3.486 (139098.93480) [0.00000]	-4.602
b(2)	-83181.7824546 (27391.48397) [0.00239]	-3.037 (29976.58061) [0.00552]	-2.775
b(3)	106.1766052 (44.40130) [0.01679]	2.391 (67.92976) [0.11804]	1.563
b(4)	525.8209297 (137.42612) [0.00013]	3.826 (177.06111) [0.00298]	2.970
b(5)	-2372094.0233687	-9.310	-7.955

	(254782.14328)	(298185.78357)
	[0.00000]	[0.00000]
b(6)	-2439822.6332825	-2.131 -1.261
	(1145102.98683)	(1934475.71223)
	[0.03312]	[0.20723]

Effective sample size (n): 37

Variance of the residuals: 113301448052.502

Standard error of the residuals (SER): 336602.80458205

Residual sum of squares (RSS): 3512344889627.57

(Also called SSR = Sum of Squared Residuals)

Total sum of squares (TSS): 21266916946763.1

R-square: 0.8348

Adjusted R-square: 0.8082

Overall F test: $F(5,31) = 31.34$
p-value = 0.00000
Significance levels: 10% 5%
Critical values: 2.04 2.52
Conclusions: reject reject

This regression has significantly stronger results, with all variables being significant. The t-stats are significant on all of the variables, the Adjusted R-square is very strong and the overall F-test (testing the soundness of the entire equation) is strong. I am happy with the number of Degrees of Freedom at 31, striking an adequate balance between having sufficient variables for the equation, and not overloading it to drive my Adjusted R-square up. It is this equation that I concluded with.

Section 5: Final Equation

RECVIS = -2439822.633 - 640104.247(GAS) - 83181.7826(ENTRY) +
106.177(HIGHWAY) + 525.821(GDP[Lag1]) - 2372094.023(dummyvisit)

The two things that would increase the visitor numbers are the amount of paved interstate system (which for the most part has been completed, and is growing only insignificantly nowadays), and the GDP per capita, which is also growing slowly. The

limiting factors in this equation, the entry fees and the gas prices are also heavy hitters - a small increase results in a significant decline in numbers (with every rise of \$1 for fuel (in 2006 terms) results in a decline of 640,000 for RECVIS). There is obviously problems with the accuracy of the modeling, and with the dataset from which the regression pulls its information, but these are statistically significant results and these variables are shown to matter.

Visitation to ANP appears to be influenced by economic variables rather than by the ambient variables such as weather with gasoline prices playing a large part in the decision to visit ANP. To give an indication of their impact according to the estimated equation, below is what predicted numbers would be for 2008 given the price of gas (still tied to 2006 dollars), all other variables held constant.

Price of Gas	number of predicted visitors
\$6.00	200,132
\$5.50	520,184
\$5.00	840,236
\$4.50	1,160,288
\$4.00	1,480,341
\$3.90	1,544,351
\$3.80	1,608,361
\$3.70	1,672,372
\$3.60	1,736,382
\$3.50	1,800,393
\$3.40	1,864,403
\$3.30	1,928,414
\$3.20	1,992,424
\$3.10	2,056,434
\$3.00	2,120,445

While these may seem extreme and unrealistic, such as having only 200,000 visitors if gas was at \$6 a gallon, I stress that its simply a prediction given past circumstances. I do not foresee there being such a drastic drop in park visits

immediately, but its foreseeable that these sort of numbers could occur if gas prices stay as they are (\$4.06 a gallon on the island) or increase over the long term.

Future research on this topic could include expanding the sample size, by finding data that allowed more years to be included. One of the chief reasons I wasnt able to use a timeseries regression was due to a lack in data for the weather variables; having that data would have strengthened my case more because I could have used the more appropriate model. I could further expand on the topic by doing a park by park comparison with similar National Parks, and watching for correlation.

For this study however, I am relatively confident that I have identified an appropriate number of the key variables that drive park visitation. With record rises in fuel prices in the last 12 months, and a struggling economy as we go into the 2008 summer season, it will be interesting to see how ANP will fare this year and in those to come.

BIBLIOGRAPHY

"Bangor International Airport." National Climatic Data Center. 4 June 2008 <<http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwDI~StnSrch~StnID~20009796>>.

CLIMAS Update - News From the Climate Assessment for the Southwest Project. Vers. Volume 8, Number 2, November 2005. The University of Arizona. 4 June 2008 <<http://www.climas.arizona.edu/pubs/update/nov2005.pdf>>.

Factor, Seth, comp. Effects of Per-Vehicle Entrance Fees on U.S. National Park Visitation Rates. 31 Aug. 2007. Nicholas School of the Environment and Earth Sciences, Duke University. 4 June 2008 <<http://hdl.handle.net/10161/397>>.

"Historical Gasoline Pump Prices (USA) - Average, Adjusted to 2006 Dollars." Swivel.Com. 4 June 2008 <http://www.swivel.com/data_columns/show/2690243>.

Jacobi, Charlie. Telephone interview. 30 Apr. 2008.

Trotter, Bill. "\$4 a Gallon? So What? State Officials Optimistic About '08 Tourism Season." Bangor Daily News. 24 May 2008. 5 June 2008 <<http://bangornews.com/news/t/news.aspx?articleid=164782&zoneid=500>>.

"USA Interstate Highway System:." Urban Transport Fact Book. 4 June 2008 <<http://www.publicpurpose.com/hwy-intmiles.htm>>.

"World Population, GDP and Per Capita GDP, 1-2003 AD (Last Update: August 2007, Copyright Angus Maddison)." 4 June 2008 <www.ggdc.net/maddison/Historical_Statistics/horizontal-file_03-2007.xls>.