

Mixed Time Series, and Cross Section Analyses of Number of Mainland U. S. Visitors to Hawaii During Years (1993-2007)

Nelson, Larry

North Carolina State University, Department of Statistics

SAS Hall, NCSU

Raleigh, North Carolina 27695

[*larry_nelson@ncsu.edu*](mailto:larry_nelson@ncsu.edu)

A visit to Hawaii during their lifetime is a dream for most citizens of the United States. The Hawaii economy has long been engaged in the tourism industry and it is one of the key industries there. The mainland United States is the principal tourism market for Hawaii. Reliable and comprehensive tourism data are collected by the Hawaii state government and are available on line. Numbers of tourists by U. S. state are reported monthly and annually so it is possible to study geographic patterns of the Hawaii market over time.

Bechdolt (1973) reported on results of cross sectional travel demand models of U. S. visitors to Hawaii (1961-1970). Vargha and Nelson (1963) also working in the cross section mode, found some of the same factors to be important determinants of tourism as reported by Bechdolt. There have been a number of statistical and computational advances since the 1960s and 1970s so it is possible to perform analyses of the refined current data but the earlier work gave good clues as to the variables to use in the models for the more recent data. Nelson, Dickey & Smith (2011) carried out a study using both a mixed time series analysis and cross sectional regressions for the years 1993 through 2007. The resulting paper emphasized the tourism management aspects, but now we are reporting on the statistical aspects of that study. It happened that there was a world-wide recession during a part of that period of time and the Hawaii market was negatively affected. One can conduct cross sectional studies every tenth year since population data are available from the United States censuses. However, for time series, one would like to use annual data. From the studies of Bechdolt (1973) and Vargha and Nelson (1963) both population of a state and average per capita income were important predictors of numbers of visitors to Hawaii. Consequently, in the present study, Gross State Product of a state on an annual basis was used since it was thought to contain both population and income components and annual data for it were available on line for each state. Another important component was airfare and airfare data were available on line for the average

annual ticket price from the principal city of each state to Honolulu, Hawaii. A sample is required from each airline of fares from one airport to another within the United States by the U. S. Bureau of Transportation Statistics and this includes fares from the principal airport of each state to Honolulu, Hawaii. These are then averaged for a given year and put on line. They were then chained to the year 2000. In keeping with economic studies of this type, the data were transformed by a logarithmic transformation (both dependent and the key independent variables). This provides elasticities such as airfare elasticity.

Model Used for Mixed Time Series Regression:

Y = Log Number of Visitors to Hawaii for a year was regressed on the following variables:

X_1 = Log of Chained Average Gross State Product

X_2 = Log of Chained Average Airfare to Honolulu

$X_{3,4, \text{and } 5}$ = Three Linear Plateau Vectors to explain variation due to three recessions and the September 11, 2001 event (after the publication of Anderson and Nelson (1975))

X_6 = Log of Distance of Principal Airport of State to Orlando, Florida

X_7 = Crude Oil Price (obtained from U. S. Department of Energy web site)

X_8 = Consumer Price Index (from U. S. Bureau of Labor Statistics web site)

X_9 = Cold Index (13 states from Northwest and North Central U. S. were given an index of 1 and all other states an index of 0)

$X_{10, \dots, 58}$ = State Effects (considered as random variables)

$X_{59, \dots, 107}$ = State x Log of Chained Average Gross State Product (considered as random variables)

Verbeke and Molenberghs (1997) discussed the use of Proc Mixed in SAS for fitting linear mixed models. They provided analysis code and result printouts. Littell et al (2000) discussed using the mixed model in situations involving repeated measures and panel studies. They emphasized the importance of studying the covariance structure in repeated measures studies. Diggle et al (2002) discussed the analysis of data from longitudinal studies with examples of biological and medical applications. The use of models in repeated measures studies was discussed in some detail.

The SAS Procedure Mixed (Littell et al, 2006) allowed the fitting of the mixed model for the 15 years x 49 states data. This procedure had been used in the medical field, but it had not been used in tourism research for studies of the type being carried out in this case. An autoregressive Model 1 error structure was modeled and the REML method was used. Denominator degrees of freedom for tests of significance were estimated using methods of Satterwaite (1946). Estimates of the random intercepts and slopes for each state in the regression of Log of Average Number of Visitors for a Given State on Log of Average Gross State Product (adjusted for other variables in the model) were measures of the interaction of State and Log of Average Gross State Product. The regression coefficients could then be graphed by state and the information used for identifying states which might need more promotional activity.

Model Used for the Cross Sectional Studies:

Y = Log of the Number of Visitors from a State to Hawaii in a year was regressed on the following variables:

X_1 = Log of Chained Average Gross State Product

X_2 = Log of Chained Average Airfare to Honolulu

X_3 = Log of Distance of Principal Airport of State to Orlando, Florida

X_4 = Cold Index

It was possible to estimate elasticities by year for both airfare and gross state product from the cross sectional studies. One could then study the pattern to see if the elasticities were changing over time. One could also look for patterns in the significance of individual terms over time to see if the relative importance of the independent factors was changing over time.

Results of Mixed Time Series Regressions

The most important fixed effects in the model were the Log Chained Average Gross State Product and one of the linear plateau variables. Log Chained Average Airfare was also significant at less than the .0001 level, but the elasticity coefficient was low implying a low airfare elasticity (-.211). All variables were significant at at least the .05 level except the Cold Index. The Log Distance to Orlando was significant and was the third most important fixed variable in the model.

Both State and State by Log Chained Average Gross State Product were significant but State was more important than the random interaction effect. A graphical presentation was made of the slopes by State and from that it was possible to recommend that the six states having slopes significantly different from zero be studied for increased promotional effort (Nevada, Utah, Arizona, South Carolina, Montana and Florida). The relative emphases to be put on these states would depend upon their populations.

Variances of the random State intercepts, the random State slopes and covariances between the intercepts and slopes are important results for the random effects. The AR (1) coefficient estimate was .398.

Results of Cross Sectional Studies

Log Chained Average Gross State Product was the most important and Log Chained Average Airfare was the second most important variable in the cross sectional studies. The Log of Orlando Distance and the Cold Index were significant in most years. The fits were quite good: R-squares of .934-.962. The airfare elasticities (median of -1.865) were much larger than that obtained in the Mixed Time Series analysis. This led to a recommendation that there should be separate terms for the two elasticities, i. e. temporal and spatial elasticities. Probably the reason for this discrepancy is that in the mixed time series model, the random State variable is highly related to distance and distance is related to airfare. So using such a model, the principal cause for airfare differences has already been removed by the State effects. The airfare differences over the 15 years were quite small so an elasticity estimate based mainly on that will not show up to be large. But in the cross section studies, the airfare differences are primarily due to distances and there are differences among the states in distances.

It was possible to relate the airfare elasticities to an index for year and this showed that there was a significant increase in elasticity over time ($\text{prob} > t = .0055$). But the airfares did not change greatly over the 15 year period so the growth in tourism over time could not be attributed to an airfare decrease over time.

Summary

A comprehensive study of tourism to Hawaii from the mainland United States for the period 1993-2007 showed that Statistics can be very useful in market research for tourism. Both mixed model time series and cross sectional models were fitted. The mixed model time series fitting was a first

for this type of application and it was made possible by SAS Proc Mixed. Two very important variables in both models were Log of Chained Average Gross State Product and Log of Chained Average Airfare. This was consistent with what was found in earlier studies of Hawaii tourism from the mainland. However, there was an inconsistency of the estimates of the airfare elasticities between the mixed time series model and the cross section model. The cross section elasticities were considerably higher and this led to the conclusion that in reporting, one should identify whether the elasticity estimates are temporal or spatial estimates.

Both types of models fit well and there were not obvious problems with the assumptions which had been made. SAS Proc Mixed provided a number of different possible categories of results for the Mixed Time Series runs and for this reason it was concluded that based on this initial application of the Mixed Time Series analysis it should become a standard in tourism research for a given destination area. But the cross section analyses should also be continued as they provide additional useful information.

In previous studies the author has found that the cross sectional studies are not respected as much by the economists as the time series studies. The present study confirms that it is very useful to perform both types of studies as the results are complementary and a better overall picture of the tourism market is attained in this way. If airlines are looking at the possible effects of changing airfares, the cross section studies would provide better estimates of elasticities at a given period of time. One can also study patterns in the elasticities over time using such a model.

The author has done tourism research in several different destinations in the Caribbean in past years. He finds that the Hawaii data are much more complete and reliable than those found in some of the Caribbean Islands. Being able to download both number of tourists or their expenditures from the web and also U. S. government records of population, per capita income, gross state product and airfares actually used during a period of time also from the web really facilitates a study of this type. Those in the tourism marketing area should concentrate more on obtaining accurate records of market performance for different destination areas.

REFERENCES

- Anderson, R. L. & Nelson, L. A. (1975). A family of models involving intersecting straight lines and concomitant experimental designs useful in evaluating response to fertilizer nutrients. *Biometrics*, 31, 303-318.
- Bechdolt, B. V. Jr. (1973). Cross-sectional travel demand functions: U. S. visitors to Hawaii, 1961-70. *The Quarterly Review of Economics and Business*, 13, 37-47.
- Diggle, P. J., Liang, K-Y, & Zeger, S. I. (2002). *Analysis of longitudinal data* (2nd ed). New York: Oxford University Press.
- Littell, R.C., Pendergrast, J., & Natarajan, R. (2000). Modeling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, 19, 1793-1819.
- Littell, R.C., Milliken, G.A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS system for mixed models* (2nd ed.), Cary, N. C: SAS Institute, Inc.
- Nelson, L. A., Dickey, D. A., & Smith, J. M. (2011). Estimating time series and cross section tourism demand models: Mainland United States to Hawaii data. *Tourism Management* 32 (1) 28-38.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110-114.
- Vargha, L. & Nelson, L. (1963). The mainland United States as a tourist market for Hawaii, Part I. A statistical analysis of geographic variation of tourist volume to Hawaii and the Caribbean. Report Prepared for the Hawaii Visitors Bureau by Statistical Research Associates of Honolulu.
- Verbeke, G. & Molenberghs, G. (1997). *Linear mixed models in practice*. New York: Springer-Verlag.