

Methods for the Analysis of Time Series of Television Audiences: A Case Study

Mario Andres Morales^{1,2}, Krishna Balasubramanian¹, Jack Smith¹, Jeffrey Storan¹ and Yuliya Torosjan¹

¹Simulmedia Inc., 670 Broadway (& Bond), 2nd Floor, New York, NY 10012

²Stanford University, Dept of Genetics, Stanford, CA 94305

Abstract

Time series were used to conduct exploratory data analysis of anonymous television viewer behaviour during the first seven episodes of the 10th season of American Idol [Jan 19 2011 to Feb 9th 2011]. A variety of visualization and graphical representation methods for this data will be compared to illustrate the observations from this analysis.

Key Words: geo experiments; television viewing; time series; exploratory data analysis; Simulmedia; marketing measurement

1. Introduction

American Idol is a music reality TV show first aired by FOX Channel in 2002 after the success of its production in the UK. During the lifetime of this show, it has maintained high ratings in the historical statistics reported by Nielsen. We focus on the graphical exploration to try to contrast inter-temporal and geographical changes.

1.1 Data Source

The data source is the Simulmedia's anonymous behavioural viewing dataset with more than 30 million US TV viewers observed second by second.

1.2 Graphics

In this observational study we examined the viewership of a set of episodes during the ten seasons of American Idol. We were interested in studying the geographic and temporal differences in viewership of these episodes, and carrying out a comparison with online data tools in subsequent studies.

The graphs illustrate geographical and temporal differences in number of viewing households. This graphical representation is a novel approach to capturing at zip code level geographical differences at multiple moments of time in a big observational dataset.

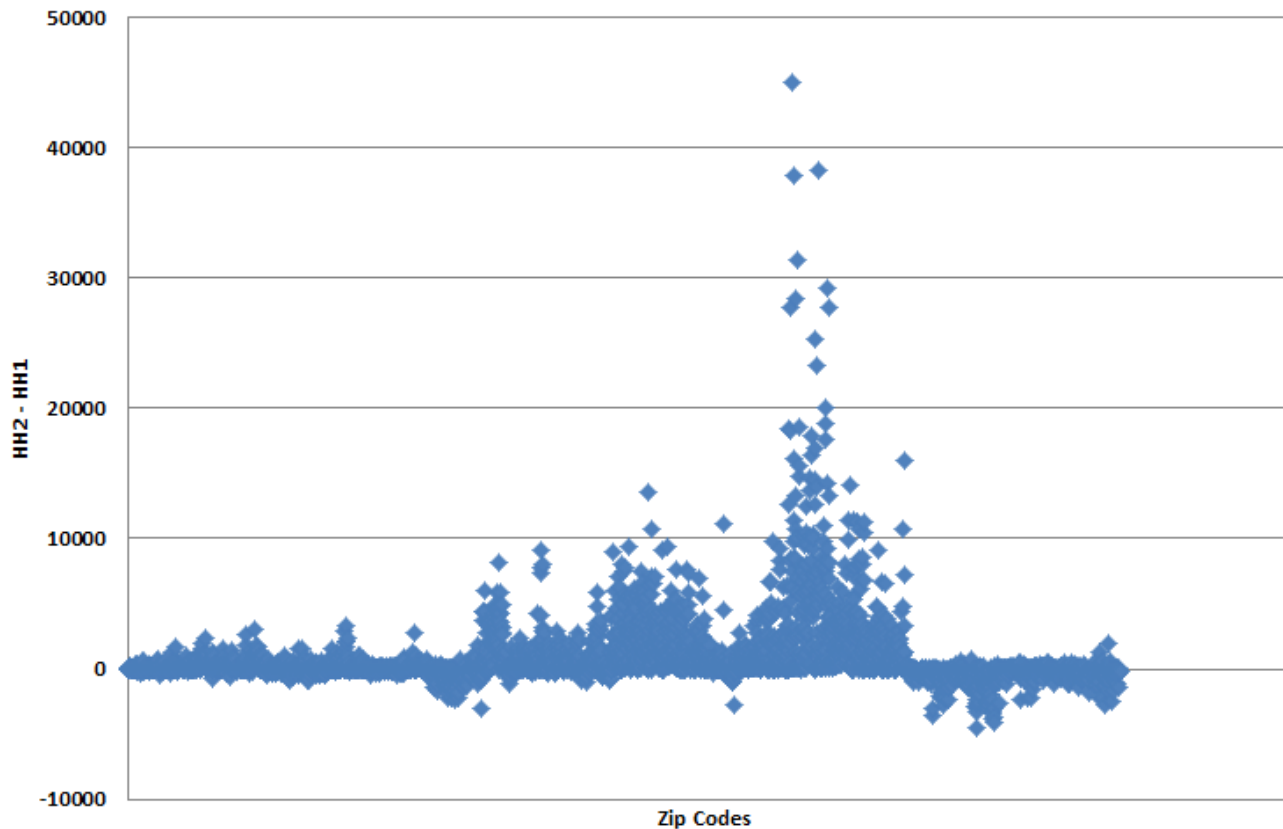


Figure 1: A delta plot. The X-axis represents a set of zip codes extracted from our sample, and the Y-axis represents the difference between episode 2 and episode 1 in the number of households (HH2-HH1) that viewed American Idol in each of the zip codes. This graph illustrates some spatial difference in the change of viewership between the two episodes of American Idol.

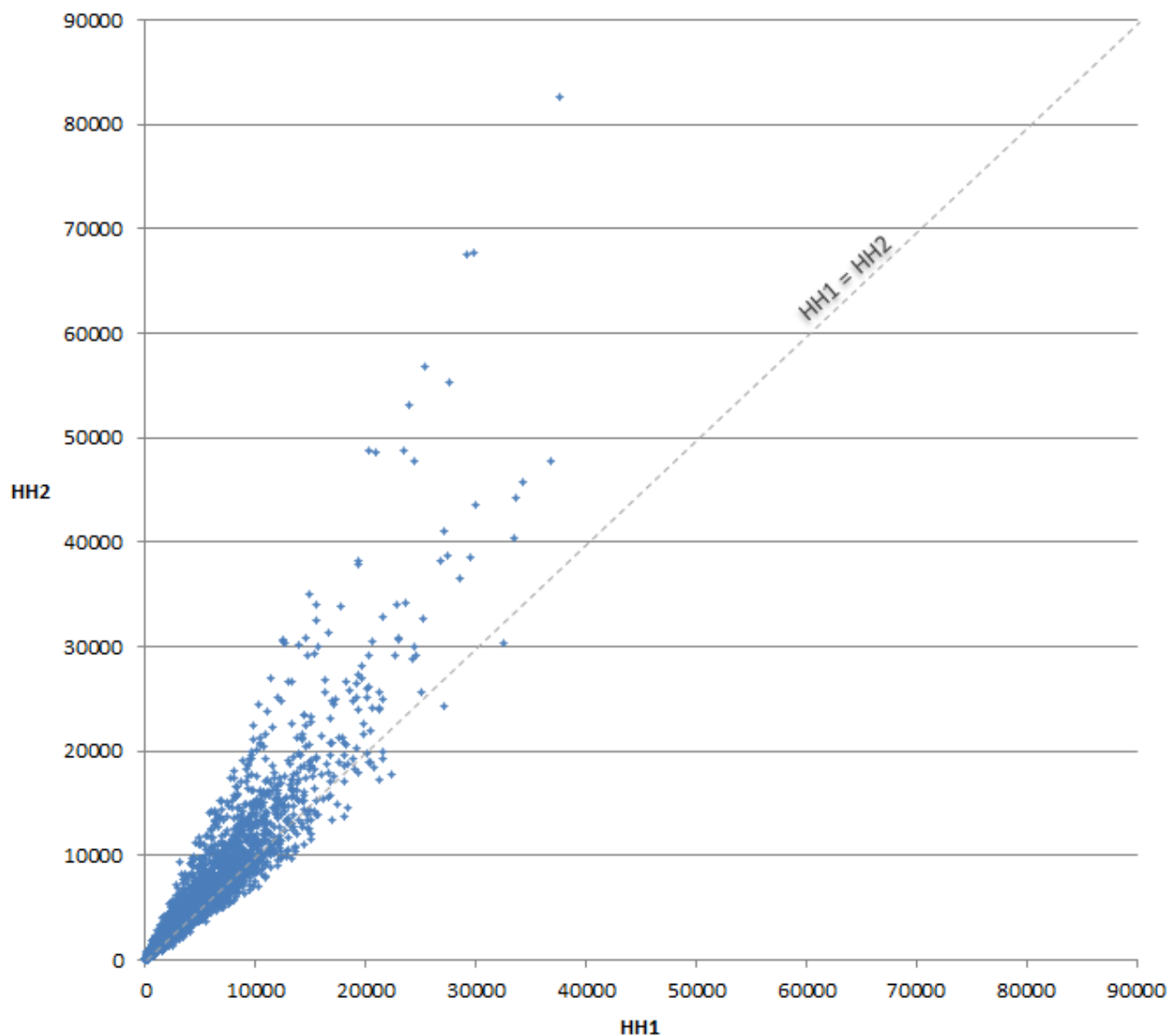


Figure 2: A scatter plot. Each point is a zip code. A reference line ($HH2=HH1$) illustrates the tendency of the viewership of the second episode (HH2) to be greater than the viewership of the first episode (HH1) for all zip codes. Additionally, this graph exhibits heteroscedasticity in the data, i.e., it shows higher variability in zip codes that have larger numbers of American Idol viewers.

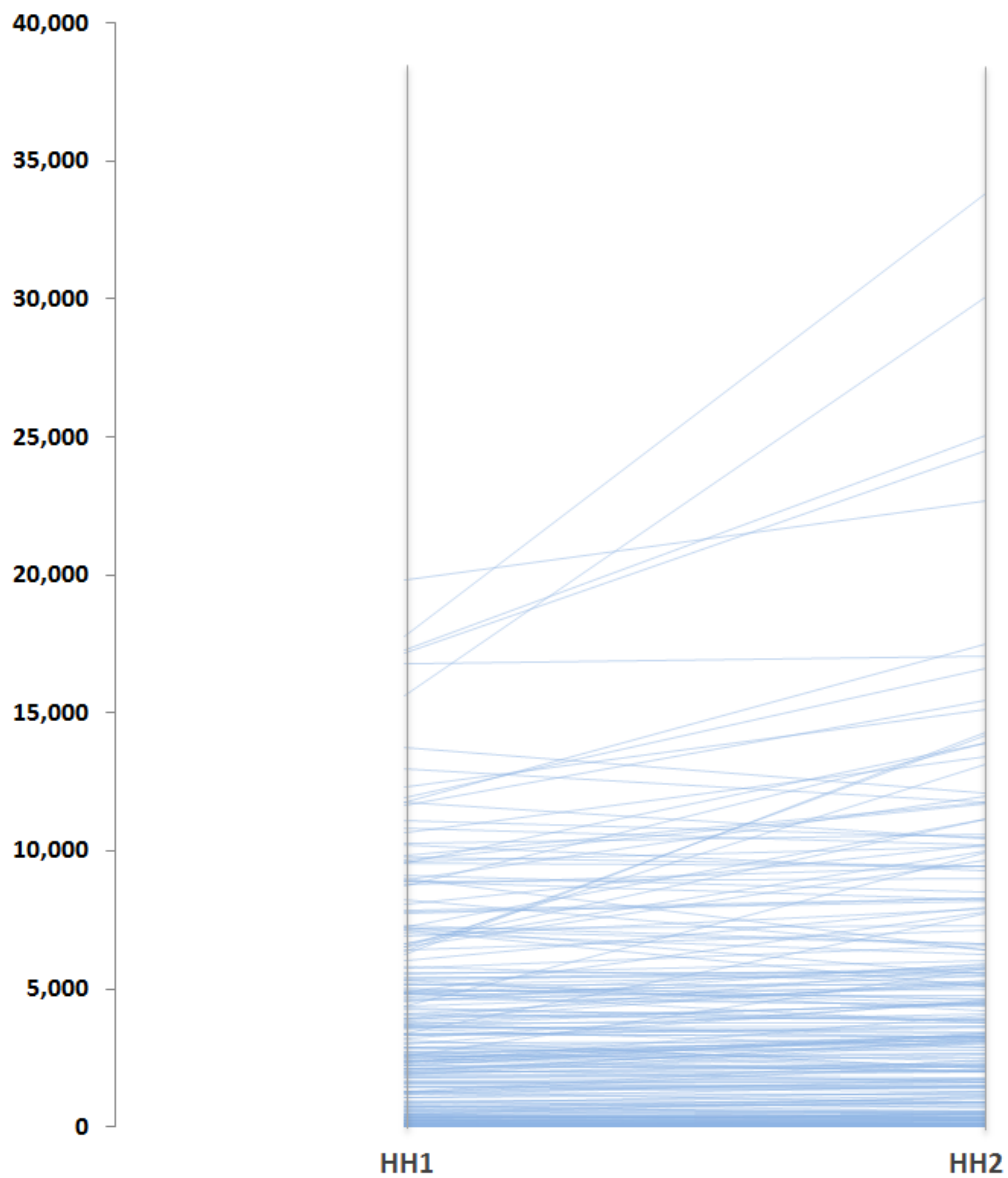


Figure 3: A stacked line chart. We took a random sample of 230 zip codes and plotted a line between the HH1 and HH2 values.

The slope of the lines reflects the change in values. The density of lines indicates the distribution of HH1 and HH2 values.

2. Conclusion

As a consequence of observing these geographical and temporal differences apparent in the graphs, we decided to use the Probabilistic Markov Chains, in order to properly model the temporal changes in different zip codes.

3. References

- Theus, M., and Urbanek, S., (2008) Interactive Graphics for Data Analysis: Principles and Examples.
- Tukey, John W.T., (1977) Exploratory Data Analysis.