Improving Bayesian methods for locating breakpoints in time series

Madison Ell[1], Dainia Higgins[2], Sarah Klingbeil[3], Nathaniel Wilson[2]

Introduction

Our project presents new approaches to finding the quantity and location of breakpoints in time series data. These methods take in account for structural changes which allows for more appropriate data modeling. Bayesian Adaptive Auto-Regression (BAAR) is a Bayesian technique that samples from a distribution of the number and location of breakpoints. New sets of breakpoints are proposed by a reversible-jump Markov Chain Monte Carlo process and are then evaluated using a Metropolis-Hasting algorithm. Simulation results show that BAAR detected the number and location of breakpoints with accuracy that met and sometimes exceeded the performance of the existing frequentist methods.

Methods

Initial Breakpoint(s): BAAR uses results from Bai-Perron[1] constrained to a maximum of 2 breakpoints combined with a burn-in period of at least 2 times the number of data points.

Step Type: Inspired by the BARS[3] technique, our Markov Chain Monte Carlo has 3 different step types: Birth, Death, and Move which encompasses 2 options called Jump and Jiggle.

- Birth: Random addition of a breakpoint.
- Death: Random deletion of a breakpoint.
- Move: Either Jump or Jiggle.

Jump: A randomly chosen existing breakpoint is moved to a random new location.

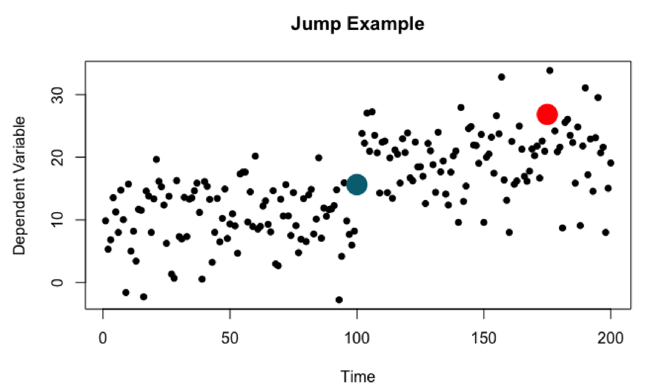


Figure 1: An example of Jump on simulated data. Existing breakpoint is in teal, and proposed is in red.

Jiggle: A randomly chosen existing breakpoint is moved to a location within a restricted interval around the original location. It is calculated by $J_n = (x_b - \rho n, x_b + \rho n)$, where n is the length of data and ρ is the percent of data in the interval.

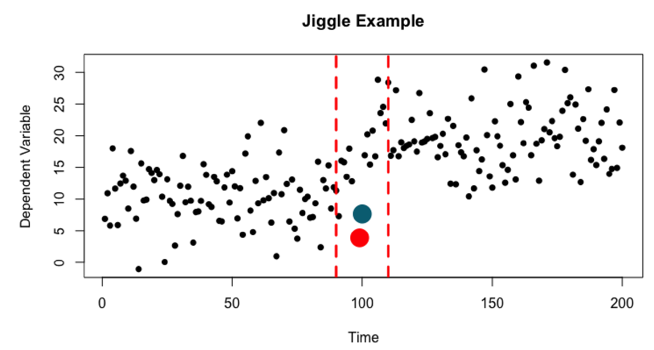


Figure 2: An example of Jiggle on simulated data with $\rho = 0.05$ and $J_n = (90, 110)$. Existing breakpoint at $t = 100$ is in teal, and proposed at $t = 99$ is in red.

The Metropolis-Hastings Ratio: Determines the acceptance of proposed breakpoints. BIC is used to approximate the Bayes factor[2], which streamlines the equation:

$$log(r) \approx \left(\frac{-\Delta BIC}{2}\right) \frac{\pi(\tau_n, K_n)}{\pi(\tau_o, K_o)} \frac{q(\tau_o K_o | \tau_n K_n)}{q(\tau_n K_n | \tau_o K_o)} \quad (1)$$

where r is the acceptance probability, K and τ are the number and location of breakpoints, q is the proposal density, and π is the prior information.

Results

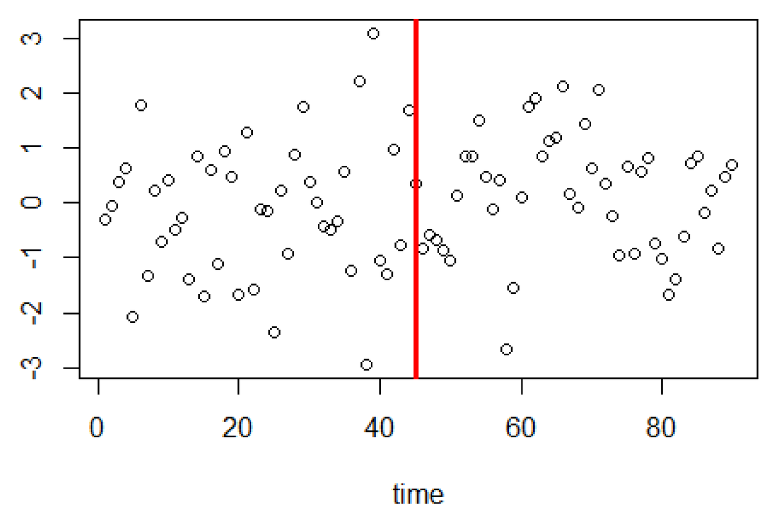


Figure 3: Simulated data used to test BAAR method with one significant break at $t = 45$. For $t \leq 45$, $y_t = -0.7y_{t-1} + \epsilon_t$ and for $t > 45$ $y_t = 0.7y_{t-1} + \epsilon_t$.

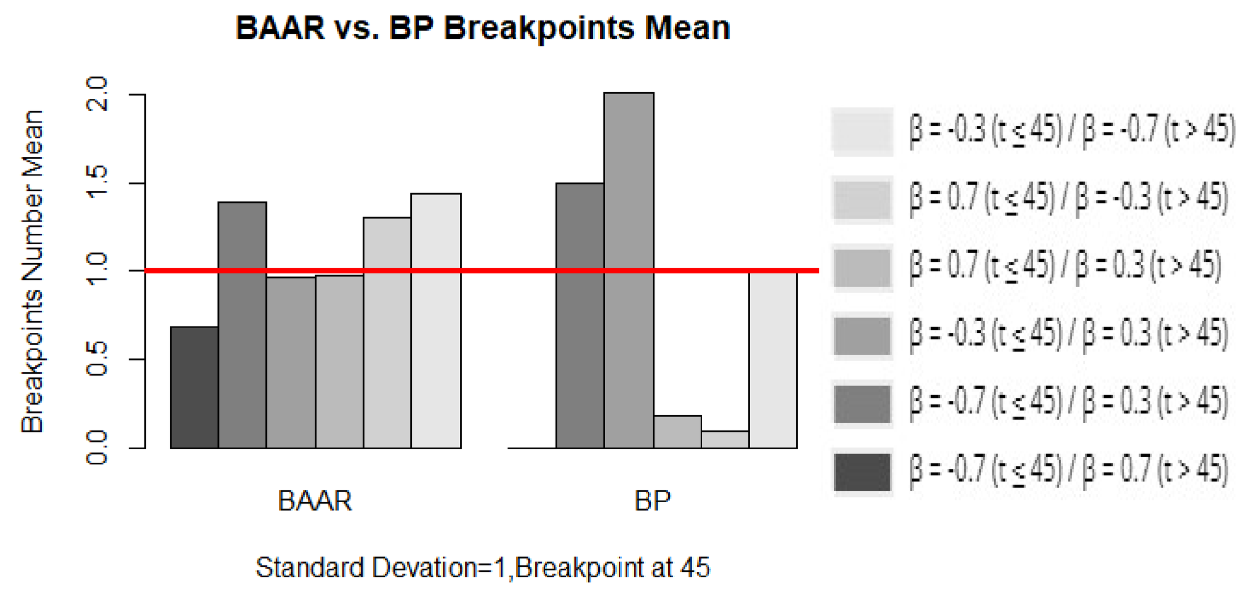


Figure 4: Average number of breakpoints obtained from a 1,000 iteration run from varying AR-1 coefficients

In the simulated data set (Figure 3), BAAR accurately finds the location of the single breakpoint (Figure 4). Due to the clarity of the breakpoint in the training data, BAAR places a high probability of it existing at exactly time point 100 while still accepting some other possibilities.

Case Study: Suicide Among People Ages 15-24 in the U.S.A.

Suicide is the fourth leading cause of death among Americans ages 15 to 24[4]. BAAR accurately models the suicide counts among people ages 15 to 24 in the United States of America as 4 subsections (Figure 5), locating breakpoints between 1995 and 1996, 2002 and 2003, and 2007 through 2013 that correspond to the economic prosperity of the 1990s, the dot-com economic crash, and the Great Recession (Figure 6).

Figure 5: Suicide counts among American people ages 15 to 24: navy blue circles - true values; orange squares - fitted values from single AR(2).

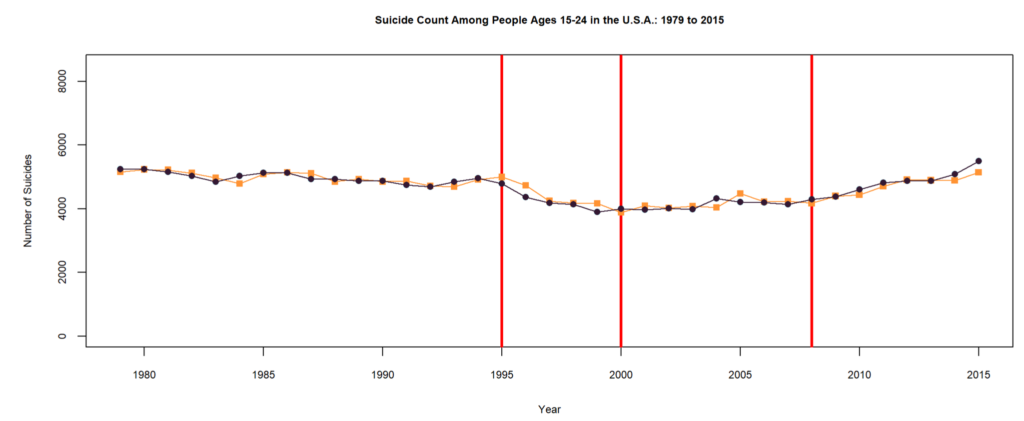
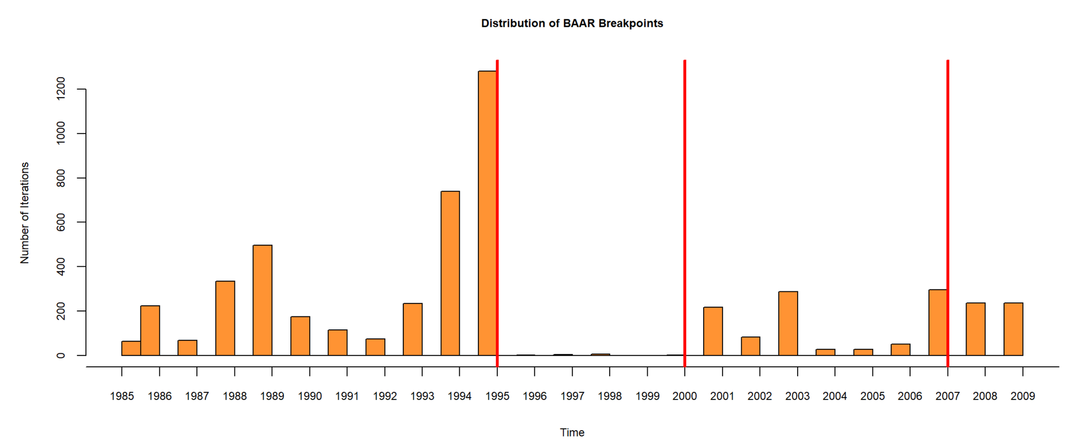


Figure 6: Distribution of breakpoint locations in suicide data among 15-24 year olds in the United States of America.



Contact Information

Emails:
madison.ell4@gmail.com
DHiggins00@student.coppin.edu
klingbeil.sarah@gmail.com
NWilson10@student.coppin.edu

References

[1] Bai, J. and Perron, P., (2003). *Computation and analysis of multiple structural change models*. Journal of applied econometrics, 18(1), pp.1-22.

[2] Kass, R.E. and Wasserman, L., (1995). *A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion*. Journal of the american statistical association, 90(431), pp.928-934.

[3] DiMatteo, I., Genovese, C.R. and Kass, R.E., (2001). *Bayesian curve-fitting with free-knot splines*. Biometrika, 88(4), pp.1055-1071.

[4] World Health Organization, (2016).*Self-inflicted injuries*. WHO Mortality Database.

Affiliations

Department of Mathematics, Lafayette College
This research is funded by the National Science Foundation, grant #2150343
[1] Mathematics and Computer Science Division, Fullerton College
[2] Department of Mathematics and Computer Science, Coppin State University
[3] Mathematics Department/English Department, Southern Adventist University