Improving Bayesian Procedures to Detect Breakpoints in Time Series Data

Madison Ell, Dainia Higgins, Sarah Klingbeil, Nathaniel Wilson
Mentor: Jeffrey Liebner

Fullerton College, Coppin State University, Southern Adventist University

Thursday, August 18, 2022

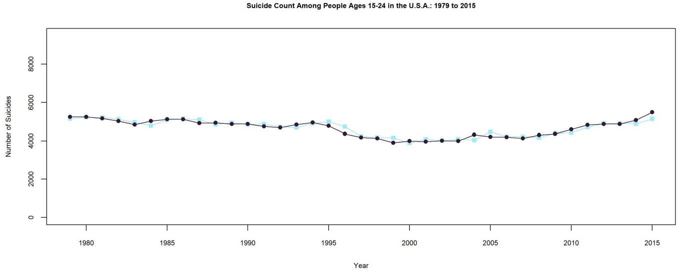
Special thanks to NSF (grant number DMS-2150343), Lafayette College and advisor Dr. Jeffrey Liebner for making this REU (Research Experience for Undergraduates) project possible.

1. Project Overview
2. Progress and Results
3. Conclusion and Further Research

What is a breakpoint?

Definition

Breakpoints (also known as change points or structural breaks) are points in which time series data changes.



How do we find breakpoints?

1. **Eyeballing**: in some cases, breakpoints are clearly identifiable due to steep trends in plotted data.

How do we find breakpoints?

1. **Eyeballing**: in some cases, breakpoints are clearly identifiable due to steep trends in plotted data.
2. **Expert Opinion**: breakpoints are approximated by experts in a specific field based on historical knowledge.

How do we find breakpoints?

1. **Eyeballing**: in some cases, breakpoints are clearly identifiable due to steep trends in plotted data.
2. **Expert Opinion**: breakpoints are approximated by experts in a specific field based on historical knowledge.
3. **Frequentist** approach: the parameter is fixed and the data is random. This is the standard type of statistics taught in introductory courses.

How do we find breakpoints?

1. **Eyeballing**: in some cases, breakpoints are clearly identifiable due to steep trends in plotted data.
2. **Expert Opinion**: breakpoints are approximated by experts in a specific field based on historical knowledge.
3. **Frequentist** approach: the parameter is fixed and the data is random. This is the standard type of statistics taught in introductory courses.
4. **Bayesian** approach: the parameter is random and the data is fixed. With this method, you apply prior information to the data that is being explored. The posterior is the result.
   - **Bayes' theorem**

$$f(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\pi(x)}$$

What is the Bai-Perron test?

Definition

The **Bai-Perron test** is a general algorithm to find an optimal breakpoint set.

1. a frequentist approach
2. checks almost every single location for a breakpoint and returns the optimal set
3. requires a user to specify the number of breakpoints

What are some common types of time series models?

Auto-regressive (AR) model: each output value depends linearly upon previous values and an independent error term

AR(1)

$x_t = \phi x_{t-1} + \epsilon_t$

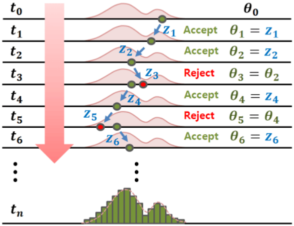
Moving average (MA) model: output value depends linearly upon previous error terms

MA(1)

$x_t = \epsilon_t + \theta \epsilon_{t-1}$

What is Markov Chain Monte Carlo (MCMC)?

1. A class of algorithms for Bayesian sampling from a probability distribution
2. Generates and records a random sample sequence from one sample to another where each proposed sample is accepted or rejected by the algorithm; this process is repeated until a stationary distribution sample is found



The Project

Bayesian Adaptive Auto-Regression (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

- needs to have an input stating breakpoint places (Bai-Perron)

The Project

Bayesian Adaptive Auto-Regression (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

- needs to have an input stating breakpoint places (Bai-Perron)
- A new breakpoint set is proposed at each step of the MCMC
  - birth, death, and move

The Project

Bayesian Adaptive Auto-Regression (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

- **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
  - Acceptance Ratio

$$ratio \approx \exp\left(\frac{-\Delta BIC}{2}\right)\frac{\pi(K_n)}{\pi(K_o)}\frac{\pi(\tau_n|K_n)}{\pi(\tau_o|K_o)}\frac{q(\tau_o K_o|\tau_n K_n)}{q(\tau_n K_n|\tau_o K_o)}$$

The Project

Bayesian Adaptive Auto-Regression (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

- **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
  - Acceptance Ratio

$$ratio \approx \exp\left(\frac{-\Delta BIC}{2}\right)\frac{\pi(K_n)}{\pi(K_o)}\frac{\pi(\tau_n|K_n)}{\pi(\tau_o|K_o)}\frac{q(\tau_o K_o|\tau_n K_n)}{q(\tau_n K_n|\tau_o K_o)}$$

- Hence, a distribution of possible breakpoints locations can be obtained

The Project

Bayesian Adaptive Moving-Average (BAMA)

BAMA is a Bayesian method used to find the location and number of breakpoints in a time series.

- Somewhat similar to BAAR in starting breakpoint and Metropolis-Hastings procedures

The Project

Bayesian Adaptive Moving-Average (BAMA)

BAMA is a Bayesian method used to find the location and number of breakpoints in a time series.

- Somewhat similar to BAAR in starting breakpoint and Metropolis-Hastings procedures
- Uses a Metropolis-Hastings procedure within a Gibbs sampler procedure to generate new coefficients

The Project

Bayesian Adaptive Moving-Average (BAMA)

BAMA is a Bayesian method used to find the location and number of breakpoints in a time series.

- Somewhat similar to BAAR in starting breakpoint and Metropolis-Hastings procedures
- Uses a Metropolis-Hastings procedure within a Gibbs sampler procedure to generate new coefficients
- Can handle seasonal/cyclical datasets

The Project

Bayesian Adaptive Moving-Average (BAMA)

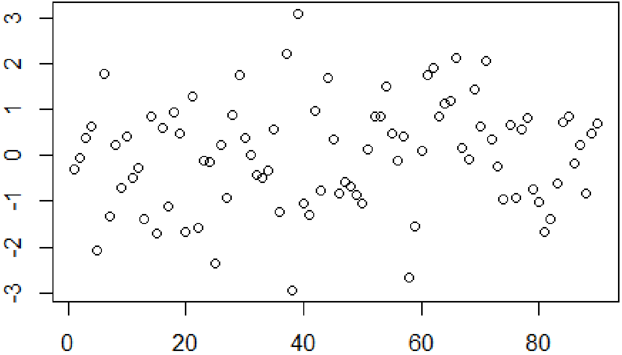
BAMA is a Bayesian method used to find the location and number of breakpoints in a time series.

- Somewhat similar to BAAR in starting breakpoint and Metropolis-Hastings procedures
- Uses a Metropolis-Hastings procedure within a Gibbs sampler procedure to generate new coefficients
- Can handle seasonal/cyclical datasets
- Improves the Bayesian adaptive algorithm compatibility with various kinds of datasets

Stress Test

Example

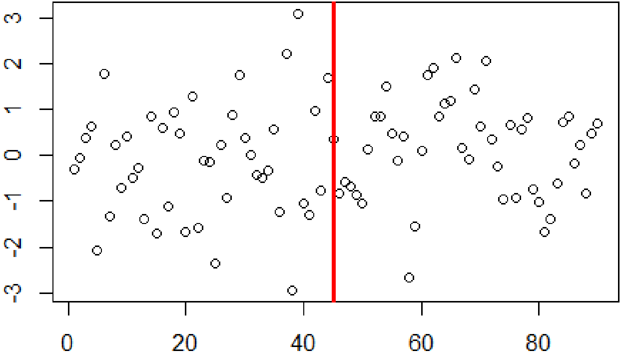
Where do you think a breakpoint is located?



Stress Test

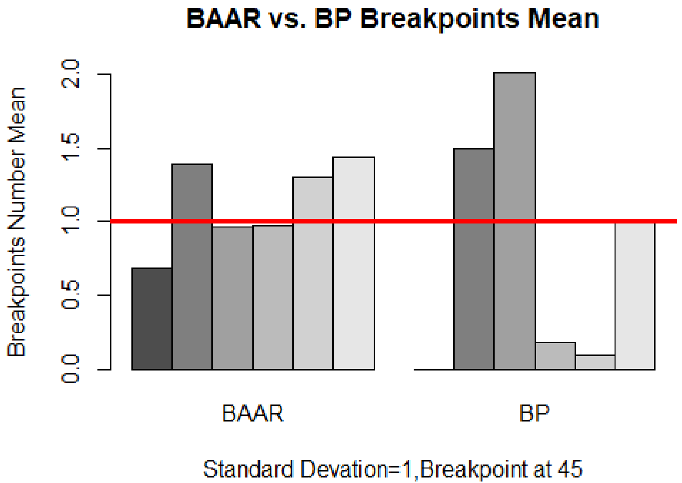
Example

How did you do?



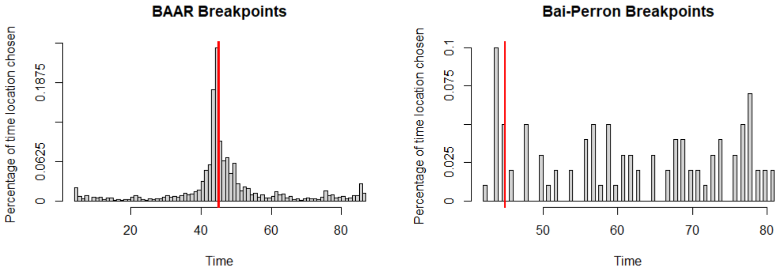
Stress Test

Simulation results

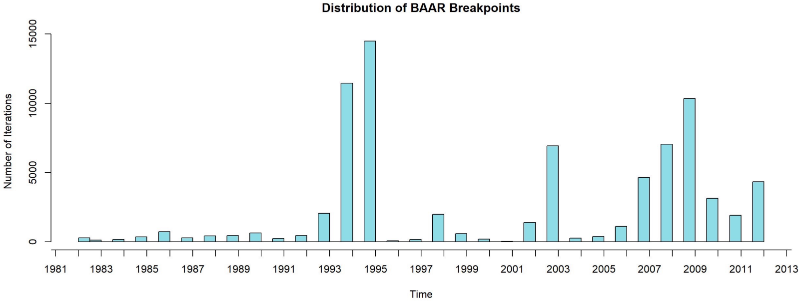


Stress Test

Simulation results

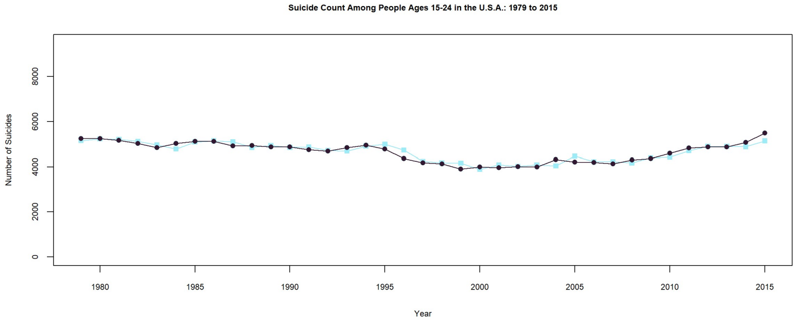


Case Study: Suicide Among People Ages 15-24 in the U.S.A.



Distribution of breakpoint locations in suicide data among 15-24 year olds in the United States of America, showing a 99% probability that a single breakpoint exists between 1994 and 1996, a 47% probability that a second breakpoint exists between 2002 and 2003, and a 73% probability that a third breakpoint exists between 2007 and 2010.

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



Suicide counts among people ages 15 to 24 in the United States of America: navy blue circles are true values; light blue squares are fitted values from a single AR(2).

1. Project Overview
2. Progress and Results
3. Conclusion and Further Research

Conclusion

- BAAR correctly identifies breakpoints with greater accuracy than existing algorithms (i.e. Bai-Perron)
- BAMA still a work in progress, in part due to COVID outbreak

Further Research

- Improve starting points of the Bayesian adaptive algorithms
- Expand BAAR and BAMA techniques to more complicated time series analysis models

References

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

Any Questions?