<u>Time series data and methods</u>

By: Patrick T. Brandt: Changepoint models

Dr. Patrick T. Brandt discusses how to look at changes of data over time to answer policy interventions. There are several review articles on state-of-the art CP models. One can analyze data over time, over geography, and want to model to see if there are statistically significant groupings over time or space. The underlying structure of this data modeling is more complicated due to changes over these variables. There are different methods to find boundaries of change points in this data. Jewell et al. (2022) focuses on both Type 1 and Type 2 models which are binary segmentation and regularization models respectively. The types of models discussed are Type 0, Type 1, Type 2, and Type 3.

Model Type 0 is a basic time series intervention model. This is used for data where there is a change at the point the analyst specifies. For this model, there is a known changepoint which is the Xt variable. For the data you would observe set of Y's and set of 's to measure change over time. You want to know the expected value of Y at different values of X. This is simple because you know when the intervention changes. This answers the question, does the change or intervention do what we expected. For this type of data, you can fit a regression, use an ARIMA model, or use a causal inference model for special cases. It is important to note that this can go wrong (Harvey and Durbin 1986) due to missing an independent variable that is not included in the model, such as further changes in the dynamics that occur after the intervention. This can create biased inferences using the Type 0 model.

Model Type 1: Binary segmentation. This is a monitoring model. The cumulative sum of square test and Chow F-tests are versions of this basic binary segmentation model (BS). For the Type 1 Model, there is a dependent variable explained by some independent variables and the parameters of the independent variables changes. The researcher wants to detect if the parameter changes so no known interaction with covariant needs to be known. For example, using a Type 1 model (Bai and Perron 1998, 2003) the data is split and the relationship after the split is different than before. This issues with this model are finding the optimal sub-samples for the regimes. Further, there needs to be a sizable sample with various time points. When $m > 2$, the problem is computationally expensive, however Bai and Perron (2003) have found a algorithm that is less expensive.

Model Type 2: Regularization and Fused Lasso methods: look at the optimal number of breaks and where these occur. This is usually best for the big data methods and find all possible optimal dates and make inferences around those. This is difficult to perform a priori. This is done by setting a cost function for the classification of a change point in splitting a sample. There is an iterative approach, looking at each changepoint one at a time (Bai and Perron 1998). There is another approach that looks simultaneously at all changepoints at once, using a penalty regression model. This second approach is great for big data. This type of model is great if there are known breaks in the data but you do not know when the breaks occur. Further, this model is great if there are multiple processes with variant dynamics in one time series or multiple time series. However, for this model it is important to consider the time and cost analysis to run this type of model.

Model Type 3: Bayesian methods. This type of model is helpful to specify a full probability model for the number and location of the changes. However, the speaker did not discuss this model type in depth. Issues with this model are that it is difficult to perform and is computationally intensive.