**Causal Inference for N-of-1 Observational Study**

**with State-Space Models**

*Mengyu Zhang*

*Dr. Linda Valeri*

*Mailman School of Public Health, Columbia University*

1. **Overview and Student Role**
2. **Background**

With advanced mobile phone technologies and accessibility to many kinds of sensors, smartphones and other wearable sensors are able to continuously collect social activity information of patients with schizophrenia (Alina Trifan, Maryse Oliveira, 2019), such as GPS, accelerometer data, call/text frequency, calling duration and survey answer. The measurements of one individual is a multivariate time series, which can be seen as N-of-1 time series studies to identify potential causal relationships. For example, we might be interested in understanding the causal relationships between call frequency behavior and perceptional loneliness. The assumptions to identify causation need to be carefully reviewed when conducting observational studies, since the presence of confounding, missing data and non-stationarity of time series would compromise the validity of estimated causal relationships.

There are many promising models to estimate the treatment effect in N-of-1 time series studies, and there are many types of treatment effect that we can estimate. The article (Eric J. Daza, 2018) uses g-formula propensity model based on a counterfactual framework to estimate average period treatment effect (APTE) for an individual. They use two ways to calculate the APTE. One is the total mean of change in outcomes which is modeled by previous outcome. The other one is the contrast of mean of change in outcome which is modeled by previous period’s last outcome or the average of its stable outcome. Shu and Peter propose Causal Transfer method to learn the effect of the treatment with state-space model (Shu Li, Peter Bu ̈hlmann, 2020) in both forms of the population or sample version, e.g., the average treatment effect (ATE), the sample average treatment effect (SATE), the conditional average treatment effect (CATE), or the marginal conditional average treatment effect (MCATE).

However, there is few articles about the application of state-space model on APTE estimate, and the link between the parameters of state-space model and APTE have not been found yet. Therefore, the goal of this project is to formulate the assumptions required to identify the APTE that is defined in (Eric J. Daza, 2018). We will use state-space model to estimate the causal effect. Also, we will formulate the causal effect of social activity (calling and texting frequency or calling duration) on clinical outcomes (mental score) in N-of-1 time series studies.

1. **Methods**
   1. **Notation**

Let and denote the exposure and outcome on period *t* time point *j*. Let denote a time point within period . represent a stochastic process. Each individual has repeated measurements at time point within time period *t*. For any random variable *B*, let . denotes association between non-causal covariates and outcome. denotes causal effect between exposure and outcome.

Let represents the potential outcome (PO) when. The point treatment effect is defined as a contrast between and . Point treatment only consists of one time point, otherwise it is period treatment. Consequently, APTE during peiord *t*, , is specified with the contrast of and *.*

* 1. **Assumptions**

To identify causal relationship, assumptions are indispensable. The first assumption is of importance called period-stable. We define the association of an outcome with a predictor as period stable if their associations at are identical for any pair of points at any *t*, i.e., and,which implies that and , or simply where .

Temporality requires that proceeds . Causal consistency ensures that the outcome we observe is identical to its corresponding potential outcome, i.e., . Exchangeability holds when potential outcome not depends on the treatment assignment, i.e., . Conditional exchangeability is that given all other covariates the independence holds, i.e., . Positivity states that for every set of values of other covariates, treatment assignment was not deterministic, i.e. , for all s.t. . SUTVA will be to extend to period level, since we were considering APTE. That means consider each stable period as a unit, therefore, there will be no inference among periods, and for each period there is one version of exposure.

* 1. **Data generation**

Recall the context of our study. We are estimating the causal effect between call frequency and mental health score that reflect perceptional loneliness. The outcome was defined as mental health score. The exposure at a certain time point was defined as calling frequency that greater than times or not, coded as 1 and 0 respectively. The constructed exposure would be the average days that call frequency is great than times in a period that is period-stable.

The data-generating process (DGP) for raw exposure would be

1. **Results**
2. **Conclusions/Discussion**
3. **References**

Trifan, A., Oliveira, M., & Oliveira, J. L. (2019). Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. *JMIR mHealth and uHealth*, *7*(8), e12649.

Daza, Eric J. "Causal analysis of self-tracked time series data using a counterfactual framework for N-of-1 trials." *Methods of information in medicine* 57.S 01 (2018): e10-e21.

Li, Shu, and Peter Bühlmann. "Estimating heterogeneous treatment effects in nonstationary time series with state-space models." *arXiv preprint arXiv:1812.04063* (2018).

Imai, Kosuke, In Song Kim, and Erik Wang. "Matching methods for causal inference with time-series cross-section data." *Princeton University* 1 (2018).

Daza, Eric Jay. "Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Tracked Health Data." *arXiv preprint arXiv:1901.03423* (2019).