Proposal

# **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 activity 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.

# **Methods**

## **Causal Effect of Interest**

“An n-of-1 study is a structured time series of outcomes, where the structure is a partition of the outcome time series imposed by a specified series of treatment periods.” (E. J. Daza, 2019).

Before defining the causal effect, we assume that association between an outcome and a predictor is stable and period stable (Eric J. Daza, 2018) so that the time series are stationary (weak stationary when there exists autocorrelation).

Let and denote the treatment and outcome on period *t* time point *j*. Treatment period *t* including a set of measurement time points . Let represents the potential outcome (PO) when . The *point treatment effect* is defined as a contrast between and . The period of *point treatment* only consists of one time point, otherwise it is *period treatment*. Therefore, the period treatment effect (PTE) is defined as the ordered set of point treatment effect. Consequently, APTE at point *j* (APTE*j*) is specified with and . With stable and period stable assumptions, APTE can be specified with and .

Daza (2019) mentioned that there are three assumptions to identify statistical causal relationships. (1) Causal Consistency (CC) ensures that the outcome we observe is identical to its corresponding potential outcome, i.e., . (2) Exchangeability holds when potential outcome not depends on the treatment assignment, i.e., . Conditional exchangeability is that given all other causes the independence holds, i.e. . (3) Positivity is also required for performing estimation. It states that for every set of values of other causes, treatment assignment was not deterministic.

## **Scenarios**

There are many kinds of models could be defined. (1) Considering the confounders. (2) Whether the exposure from previous period would affect the outcome in current period, i.e., considering carryover causal effect. (3) Whether the outcome from previous period would affect the outcome in current period, i.e., considering autocorrelation in outcome.

## **Data and Generation**

We assume that we have period exposures instead of treatments, i.e., for. Besides binary exposure {(*X*)} (high or low social activity) and continuous outcome {(*Y*)} (mental score), predictors of treatment {(*Z*)} and simultaneous causes {(*V*)} will also be generated. {(Z)} and {(*V*)} are stationary. For any random variable *B*, let  .

First, we consider the simplest model where there is no variable that confounding the causal relationships between exposure and outcome. The data generation process (DGP) for *X* would be , for , where. Data generation model (DGM) could use

where . DGP for outcome would be

where represents the carryover effect from previous periods, and represents the autocorrelation in {(*Y*)}. The DGM could be state-space model

This model can be used to generate non-stationary outcome when the other variables are non-stationary too. If we want to generate stationary outcome, we can do some changes as following.

can be stationary in period *t* if we change the its state equation to

or AR (1) process

with the expectation and variance . can be stationary in period *t* if it is defined as

or AR (1) process

with expectation and variance . The variance of and are all independent for any *i*, *t* and *j*.

When all state equations and {(*X*)} from periods to period have a stationary process, outcome would be stationary, with expectation and variance

Second, when confounders are present, DGP for *Y* stays the same, while DGP for *X* is more complex and can be , where . Let , and let   represents , so would be a set of confounders. DGM is similar, , and the probability would be calculated by logistic model.

Besides state-space model, we will use other DGM of outcome… ARMA…

## **Estimation of Causal Effect**

First, conduct changepoint analysis on the exposure series…

## **Assumption Violations**

After the estimation, simulations will be conducted to assess how missing data/non-stationarity of the time series leads to violation of the assumptions. The violation will be evaluated from several aspects, bias of the estimators, confidence interval of estimators and hypothesis tests…

# **Timeline**

|  |  |  |
| --- | --- | --- |
| Phase | Period | Task |
| Phase one | June 6th to July 31st | Literature Research |
| Phase two | July 6th to July 31st | Data simulation |
| Phase three | Aug 1st to Oct 31st | Data Analysis |

Reference:

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).