Causal Inference for N-of-1 Observational Study with State-Space Models

Mengyu Zhang Dr. Linda Valeri*

1 Overview and Student Role

The project “Causal Inference for N-of-1 Observational Study (N1OS) with State-Space Models” is supervised by Dr. Linda Valeri in Biostatistics Department at Columbia University. My responsibility is to find the connection between parameters in state space model and treatment effect in N1OS by simulations.

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. Multivariate time series data are naturally generated when a person is monitored over time (Scott L. Zeger et al, 2006), which can be seen as N-of-1 time series studies to identify potential causal relationships. 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.

Many methods have been developed for time series causality in different domains such as policy (Alberto Abadie, 2010), economics (Alberto Abadie, 2003; Susan Athey, 2017), biomedical informatics (Samantha Kleinberg, 2011). Granger causality is a useful framework for inference, since Granger and structural causality are equivalent under assumption of conditional exogeneity (White and Kennedy, 2009; White and Lu, 2010). For N1OS, Eric J. Daza (2018) uses g-formula propensity model based on a counterfactual framework to estimate average period treatment effect (APTE) for an individual. 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.

*Supervisor, Columbia University

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 text frequency or calling duration) on clinical outcomes (mental score) in N1OS.

3 Methods

3.1 Notations

Let $X_{t(j)}$ and $Y_{t(j)}$ denote the exposure and outcome on period t time point j. Let $t(j)$ denote a time point within period $t \in (1, \ldots, \tau)$. $\{(X, Y)\}$ represent a stochastic process. Each individual has repeated measurements at time point $j \in (1, \ldots, m_t)$ within time period t. For any random variable B, let $\bar{B}_{t(j)} = \left(B_{t(j)}, B_{t(j-1)}, \ldots, B_{t(1)}, B_{t-1(m_{t-1})}, B_{t-1(m_{t-1}-1)}, \ldots\right).\beta_{t(j)}$ denotes association between non-causal covariates $U_{t(j)}$ and outcome. $\mu_{t(j)}$ denotes causal effect between exposure and outcome.

Let Y_t^a represents the potential outcome (PO) when $X_t = a$. The point treatment effect is defined as a contrast between $Y_{t(j)}^a$ and $Y_{t(j)}^{a'}$. Point treatment only consists of one time point, otherwise it is period treatment. Consequently, APTE during peiord t, apte $_t$, is specified with the contrast of $E\left(Y_{t(j)}^a\right)$ and $E\left(Y_{t(j)}^{a'}\right)$

3.2 Assumptions

A key assumption to identify causal relationship in N1OS is period-stable. We define the association of an outcome with a predictor as period stable if their associations at $\{t, t' \neq t\}$ are identical for any pair of points at any t or from the same normal distribution, i.e., $\beta_{t(j)} \sim N(\beta_t^0, \sigma_{w_t}^2)$ and $\mu_{i,t(j)} \sim N(\mu_{i,t}^0, \sigma_{i,u_t}^2)$.

Temporality requires $X_{t(j)}$ proceeds $Y_{t(j)}$. Causal consistency ensures that the outcome we observe is identical to its corresponding potential outcome, i.e. $Y_{t(j)} = \sum_a Y_{t(j)}^a I\left(X_{t(j)} = a\right)$. Exchangeability holds when potential outcome not depends on the treatment assignment, i.e. $\left\{Y_{t(j)}^a\right\} \perp X_{t(j)}$. Conditional exchangeability is that given all other covariates the independence holds, i.e., $\left\{Y_{t(j)}^a\right\} \perp X_{t(j)} \mid U_{t(j)}$. Positivity states that for every set of values of other covariates, treatment assignment was not deterministic, i.e. $\Pr\left(X_{t(j)} = a \mid U = u\right) > 0$, for all U s.t. $\Pr(U = u) \neq 0$. SUTVA will be to extend to period level, which means each stable period is considered as a unit. Therefore, there will be no interference among periods, and for each period there is one version of exposure.

3.3 Estimate Causal effect

Definition of APTE

Let $\alpha_{t(j)}$ represent the values of the state at time point $t(j)$. The outcome is denoted by a vector $y_{t(j)}$ which is a linear combination of treatment and other covariates $Z_{t(j)}$ with coefficient matrix $\alpha_{t(j)}$. In time period t the state space model is

$$\begin{aligned} y_{t(j)} &= Z_{t(j)}\alpha_{t(j)} + \varepsilon_{t(j)} \\ \alpha_{t+1(j)} &= T_{t(j)}\alpha_{t(j)} + R_{t(j)}\eta_{t(j)} \end{aligned} \tag{1}$$

where $\alpha_{t(1)} \sim N(a_{t,1}, P_{t,1})$, $\varepsilon_{t(j)} \sim N(0, H_{t(j)})$, and $\eta_{t(j)} \sim N(0, Q_{t(j)})$. Random variation has mean zero and is uncorrelated over time.

In state space model the estimation of α was based on Kalman Filter. The following discussion is in a specific time period, so for simplicity, we change $t(j)$ to j. Let $a_{j|j} = E(\alpha_j|\bar{Y}_j)$, $a_{j+1} = E(\alpha_{j+1}|\bar{Y}_j)$, $P_{j|j} = Var(\alpha_j|\bar{Y}_j)$ and $P_{j+1} = Var(\alpha_{j+1}|\bar{Y}_j)$. Therefore, the filtering distributions of α_j given Y_j and α_{j+1} given $\bar{Y}_j$ are given by

$$\alpha_j|\bar{Y}_j \sim N(a_{j|j}, P_{j|j}) \tag{2}$$

and

$$\alpha_{j+1}|\bar{Y}_j \sim N(a_{j+1}, P_{j+1}) \tag{3}$$

The Kalman filter is a set of recursion equations for determining the optimal estimates of the state vector given information available at a certain time. The recursion equations are

$$\begin{aligned} v_j &= y_j - Z_j a_j \text{ (forecast error)} \\ F_j &= Z_j P_j Z_j' \\ a_{j|j} &= a_j + P_j Z_j' F_j^{-1} v_j \\ P_{j|j} &= P_j - P_j Z_j' F_j^{-1} Z_j P_j \\ a_{j+1} &= T_j a_j + K_j v_j \\ P_{t+1} &= T_j P_j (T_j - K_j Z_j)' + R_j Q_j R_j' \end{aligned} \tag{4}$$

for $t = 1, ..., m_t$, where *Kalman gain* $K_t = T_t P_t Z_t' F_t^{-1}$.

Model

Since Xs are genereate completely at random, causation between X and Y will not be confounded by other variables. Therefore the model we can used for estimation is

$$\begin{aligned} E(Y_{t(j)}) &= \beta_{0,t} + \beta_{1,t(j)}V_{t(j)} + \mu_{1,t(j)}X_{t(j)} \\ \beta_{1,t(j+1)} &= \beta^0_{1,t} \\ \mu_{1,t(j+1)} &= \mu^0_{1,t} \end{aligned} \tag{5}$$

The model of mental score against lagged scores and lagged exposures are also estimated and defined as following.

$$\begin{aligned} E(Y_{t(j)}) &= \beta_{0,t} + \beta_{1,t(j)}V_{t(j)} + \mu_{1,t(j)}X_{t(j)} + \mu_{2,t(j)}X_{t(j-1)} + \varphi_{1,t}Y_{t(j-1)} + \varphi_{2,t}Y_{t(j-2)} \\ \beta_{1,t(j+1)} &= \beta^0_{1,t} \\ \mu_{1,t(j+1)} &= \mu^0_{1,t} \\ \mu_{2,t(j+1)} &= \mu^0_{2,t} \\ \varphi_{1,t(j+1)} &= \varphi_{1,t(j)} \\ \varphi_{2,t(j+1)} &= \varphi_{2,t(j)} \end{aligned} \tag{6}$$

$\mu_{i,t(j)}$, $a_{i,t(j)}$ from the filtering distribution of $\mu_{i,t(j)}|\bar{Y}_{t(j-1)} \sim N(a_{i,t(j)}, P_{i,t(j)})$ are parameters of interests. We define immediate APTE under certain calling frequency level as numerical average coefficient of $X_{t(j)}$ within a certain time period t.

$$\begin{aligned} apte_t &= E\left(E\left[Y^{\boldsymbol{b_{t(j)}}}_{t(j)} \mid X_{t(j-1)}, \bar{Y}_{t(j-1)}\right] - E\left[{Y^{\boldsymbol{b_{t(j)}}}_{t(j)}}^* \mid X_{t(j-1)}, \bar{Y}_{t(j-1)}\right]\right) \\ &= E(\mu_{1,t(j)}) \\ &= 1/m_t \sum_j \mu_{1,t(j)} \end{aligned} \tag{7}$$

where $\boldsymbol{b_{t(j)}} = b_{t(j-1)}, b_{t(j)}$. ${Y^{\boldsymbol{b_{t(j)}}}_{t(j)}}^* | \boldsymbol{X_{t(j)}} = \boldsymbol{b^*_{t(j)}}$ is the counterfactual of $Y^{\boldsymbol{b_{t(j)}}}_{t(j)} | \boldsymbol{X_{t(j)}} = \boldsymbol{b_{t(j)}}$. Estimator would be

$$\begin{aligned} \hat{apte}_t &= 1/m_t \sum_j \hat{\mu}_{1,t(j)} \\ &= 1/m_t \sum_j a_{1,t(j)}, \text{ based on filtering distribution (3)} \end{aligned}$$

Similarly, lagged APTE under certain calling frequency level is defined as as numerical average coefficient of $X_{t(j-1)}$ within a certain time period t.

$$
\begin{aligned}
apte_t^- &= E\left(E\left[Y_{t(j)}^{\boldsymbol{b_{t(j)}}} \mid X_{t(j)}, \bar{Y}_{t(j-1)}\right] - E\left[Y_{t(j)}^{\boldsymbol{b_{t(j)}}^*} \mid X_{t(j)}, \bar{Y}_{t(j-1)}\right]\right) \\
&= E(\mu_{2,t(j)}) \\
&= 1/m_t \sum_j \mu_{2,t(j)}
\end{aligned} \tag{8}
$$

Estimator would be

$$
\begin{aligned}
\hat{apte}_t^- &= 1/m_t \sum_j \hat{\mu}_{2,t(j)} \\
&= 1/m_t \sum_j a_{2,t(j)}, \text{ based on filtering distribution (3)}
\end{aligned}
$$

Estimation

Steps for APTE estimation are

1. **Overall fitting**. Using state space model to fit whole time series and detect the change point of estimated immediate APTE.
2. **Periods separation**. Based on the change point of estimated APTE, separated whole series into several periods that is period stable.
3. **Separately fitting**. When fitting whole time series, state space model uses information from all previous points, which violated the assumption SUTVA. Therefore, models are fitted separately, and $\mu_{i,t(j)}$ is obtained for each stable period t.
4. **APTE estimation**. Finally, calculate $\hat{apte}_t$ or $\hat{apte}_t^-$.

3.4 Data generation and Simulation

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. We call one day is high call-frequent when call frequency is greater than f times. Therefore, the exposure in period t at time point j $X_{t(j)}$ was coded as 1 when day is high call-frequent and 0 when it is low call-frequent. The constructed exposure $X_t = \frac{1}{m_t}\sum_{i=1}^{m_t} X_{t(j)}$ would be the proportions of high call-frequent days, which is also a parameter used to generate exposure. Besides binary exposure $\{(X)\}$ (high or low social activity) and continuous outcome $\{(Y)\}$ (mental score), simultaneous causes $\{(V)\}$ will also be generated where $\{(V)\}$ is stationary.

The data-generating process (DGP) for raw exposure would be $X_{t(j)} = g_{t(j)}^X(X_{t(j)} = a, \xi)$, for $j \in (1, \ldots, m_t)$, where $\xi \sim Uniform(0,1)$. DGM could be

$$X_{t(j)} = I(\xi \leq Pr(X_{t(j)} = 1)), \tag{9}$$

DGP for outcome is

$$Y_{t(j)} = g^{Y}_{t(j+1)}(X_{t(j)}, \bar{X}_{t(j-1)}, \bar{Y}_{t(j-1)}, V_{t(j)}), \tag{10}$$

Data-generating model (DGM) for outcome is

$$\begin{aligned} Y_{t(j)} = \beta_{0,t} + \beta_{1,t(j)}V_{t(j)} + \mu_{1,t(j)}X_{t(j)} + \mu_{2,t(j)}X_{t(j-1)} + \varphi_{1,t}Y_{t(j-1)} + \varphi_{2,t}Y_{t(j-2)} + k_{t(j)},\ k_{t(j)} &\sim N(0, \sigma^2_{k_t}) \\ \beta_{1,t(j+1)} &= \beta_{1,t(j)} \\ \mu_{1,t(j+1)} &= \mu_{1,t(j)} \\ \mu_{2,t(j+1)} &= \mu_{2,t(j)} \end{aligned} \tag{11}$$

where $\beta_{1,t(j)} = \beta^0_{1,t}$, $\mu_{1,t(j)} = \mu^0_{1,t}$, $\mu_{2,t(j)} = \mu^0_{2,t}$ and

$$V_{t(j+1)} = V^0_t + \rho_t(V_{t(j)} - V^0_t) + v_{t(j)},\ v_{t(j)} \sim N(0, \sigma^2_{v_t}), \tag{12}$$

and for stationarity

$$\begin{cases} \varphi_{1,t} + \varphi_{2,t} < 1 \\ \varphi_{2,t} - \varphi_{1,t} < 1 \\ |\varphi_{2,t}| < 1 \\ |\rho_t| < 1 \end{cases} \tag{13}$$

Based on the generating model (9) (11) (12), 100 datasets that contains 5 periods with different level of call-frequencies are generated. The parameters are listed in Appendix. Finally, 100 simulations are conducted to estimate the immediate APTE or lagged APTE defined in session 3.3.

4 Results

Table 1 and Table 2 shows the mean and variance of immediate APTE $\hat{apte}_t$ and lagged APTE $\hat{apte}^-_t$ of model (6) for 100 simulations. Most of the estimation for five periods are close to truth except for lagged APTE in first period, which might due to the long burn-out period.

Table 1: Immediate APTE in model (6) for 100 Simulations

	Length	m_t	Estimated	True	Bias	Variance	Prop
APTE1	512.57	500	1.948	2.0	-0.052	0.041	0.857
APTE2	759.16	800	0.554	0.5	0.054	0.094	0.286
APTE3	447.57	400	1.441	1.5	-0.059	0.077	0.714
APTE4	618.10	600	0.962	1.0	-0.038	0.041	0.429
APTE5	667.60	700	0.453	0.5	-0.047	0.010	0.143

Table 2: Lagged APTE in model (6) for 100 Simulations

	Length	m_t	Estimated	True	Bias	Variance	Prop
APTE_1	512.57	500	1.286	1.5	-0.214	2.041	0.857
APTE_2	759.16	800	0.270	0.2	0.070	0.081	0.286
APTE_3	447.57	400	0.933	1.0	-0.067	0.054	0.714
APTE_4	618.10	600	0.500	0.5	0.000	0.065	0.429
APTE_5	667.60	700	0.980	1.0	-0.020	0.017	0.143

Table 3 shows the mean and variance of immediate APTE $\hat{apte}_t$ of model (5) for 100 simulations. The estimation has low variance and bias without considering lagged outcomes and exposures. A better estimation of APTE is obtained in first period comparing model (5) to model (6).

Table 3: Lagged APTE in model (5) for 100 Simulations

	Length	m_t	Estimated	True	Bias	Variance	Prop
APTE1	526.50	500	1.969	2.0	-0.031	0.032	0.857
APTE2	787.12	800	0.491	0.5	-0.009	0.006	0.286
APTE3	404.36	400	1.502	1.5	0.002	0.023	0.714
APTE4	618.83	600	0.950	1.0	-0.050	0.010	0.429
APTE5	668.19	700	0.435	0.5	-0.065	0.011	0.143

5 Conclusions and Discussion

The goal of the project is treatment effect estimation in N of 1 observational study. To validate SUTVA assumption in causal inference, the concept of period-stable, proposed by Eric Daza (2018), was introduced. Based on the period-stable assumption, daily exposure information during a stationary period was summarized to constructed exposure. Given a specific model with one lag of exposure and lag two autocorrelation of outcome, immediate APTE and lagged APTE was defined. They are estimated by filtering distribution of state vectors/coefficients vector within each period-stable period so that SUTVA assumption holds.

Although estimation results shown above is very close to the truth, real application of state space model is still questionable. The relationship between exposure and outcome could be more complex and noisier. Period-stable assumption could be violated when analyzing subjects with unstable mental status. Effect carryover between adjacent periods and effect slow onset/decay are not considered in this report. Eric Daza (2018) introduced indicator variable into g-formula to accommodate effect carryover from previous treatment periods, and slow onset or decay of the effect. Indicator variable could be a promising solution for state space model as well.

Reference

Abadie, Alberto, and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country." American economic review 93.1 (2003): 113-132.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." Journal of the American statistical Association 105.490 (2010): 493-505.

Athey, Susan, and Guido W. Imbens. "The state of applied econometrics: Causality and policy evaluation." Journal of Economic Perspectives 31.2 (2017): 3-32.

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.

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).

Kleinberg, Samantha, and George Hripcsak. "A review of causal inference for biomedical informatics." Journal of biomedical informatics 44.6 (2011): 1102-1112.

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

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.

Zeger, Scott L., Rafael Irizarry, and Roger D. Peng. "On time series analysis of public health and biomedical data." Annu. Rev. Public Health 27 (2006): 57-79.

Appendix

Parameter Setting

We have five time periods, each time periods will have m_t time points. $X_t = \frac{1}{m_t}\sum_{i=1}^{m_t} X_{t(j)}$, proportions of high call-frequent days, is also pre-specified as column Prop. $\mu_{i,t(j)}$ is the effect that we are interested in.

Table 4: Parameters for time period

t	m_t	Prop
1	500	6/7
2	800	2/7
3	400	5/7
4	600	3/7
5	700	1/7

Table 5: Parameters for variance

t	$\sigma^2_{k_t}$	$\sigma^2_{v_t}$	$\sigma^2_{w_t}$	$\sigma^2_{u_{1,t}}$	$\sigma^2_{u_{2,t}}$
1	0.1	0.2	0.15	0.15	0.15
2	0.1	0.2	0.15	0.15	0.15
3	0.1	0.2	0.15	0.15	0.15
4	0.1	0.2	0.15	0.15	0.15
5	0.1	0.2	0.15	0.15	0.15

Table 6: Parameters for mean of coefficients

t	$\beta_{0,t}$	$\beta^0_{1,t}$	$\mu^0_{1,t}$	$\mu^0_{2,t}$	V^0_t
1	1	1.5	2	1.5	1.5
2	1	2.0	0.5	0.2	1
3	1	1.5	1.5	1.0	0.8
4	1	1.0	1	0.5	1.2
5	1	0.5	0.5	0.2	0.9

Table 7: Parameters for AR model coefficients

t	$\varphi_{1,t}$	$\varphi_{2,t}$	ρ_t
1	0.1	-0.15	0.1
2	0.1	-0.1	0.15
3	0.05	-0.125	0.125

t	$\varphi_{1,t}$	$\varphi_{2,t}$	ρ_t
4	0.15	-0.125	0.125
5	0.05	-0.15	0.075