Estimation of the causal estimand (Homework 4)

Yancheng Li

November 10, 2022

The causal estimand we want to estimate is

$$E[Y_t(\overline{a}_t)] = \int \cdots \int E[Y_t|\overline{A}_t = \overline{a}_t, \overline{Y}_{t-1} = \overline{y}_{t-1}] \prod_{s=1}^{t-1} P(y_s|\overline{y}_{s-1}, \overline{a}_s) dy_s$$

However, we cannot simply plug in estimates of all the unknown quantities in the above formula (g-computation) due to dimensionality (in non-parametric estimates) and the null-paradoex (in parametric estimates). Instead, we use a semiparametric model called the marginal structural model (MSM) to do the estimation, which means we directly specify a parametric functional form from $g(\overline{a}_t, \beta)$ for $\psi(\overline{a}_t)$.

We approximate/simplify the SIR (Susceptible, Infected, Recovered) model, and propose the following MSM model:

$$E[L_t(\overline{a}_t)] = \nu(t) + \beta M(\overline{a}_t)$$

where $L_t = \log(Y_t + 1)$ and we approximate $\log E[Y_t(\overline{a}_t)]$ with $E[\log Y_t(\overline{a}_t)]$, $M(\overline{a}_t) \equiv \sum_{s=1}^{t-\delta} a_s$, and $\nu(t) = \sum_{j=1}^{k} \beta_i \psi_j(t)$. We estimate $\nu(t)$ and β by solving the following estimating equation

$$\sum_t h_t(\overline{a}_t) W_t [L_t - (\widehat{\nu}(t) + \widehat{\beta} M(\overline{a}_t))] = 0$$

where $h_t(\overline{a}_t) = (1, \psi_1(t), \cdots, \psi_k(t), M(\overline{a}_t))^T$ and

$$W_t = \prod_{s=1}^{t} \frac{\pi(A_s|\overline{A}_{s-1})}{\pi(A_s|\overline{A}_{s-1}, \overline{Y}_{s-1})}$$

We estimate the densities in the weights equation by a moment-based approach: the vector of weights $W_1, \cdots, W_T$ need to satisfy the following moment constraints:

$$E[h_1(A_t) h_2(Y_{t-1}) W_t] = 0$$

where

$$\begin{aligned} h_1(a_t) &= \widetilde{h}_1(a_t) - E[\widetilde{h}_1(A_t)|\overline{A}_{t-1}] \\ h_2(y_{t-1}) &= \widetilde{h}_2(y_{t-1}) - E[\widetilde{h}_2(y_{t-1})|\overline{A}_{t-\delta-1}, \overline{Y}_{t-2}] \end{aligned}$$

To solve this, for some k, make a Markov assumption

$$\begin{aligned} \mu &= E[\widetilde{h}_1(A_t)|\overline{A}_{t-1}] = E[\widetilde{h}_1(A_t)|A_{t-1}, \cdots, A_{t-k}] \\ \nu &= E[\widetilde{h}_2(Y_{t-1})|\overline{A}_{t-\delta-1}, \overline{Y}_{t-2}] = E[\widetilde{h}_2(Y_{t-1})|A_{t-k-\delta-1}, \cdots, A_{t-\delta-1}, Y_{t-1-k}, \cdots, Y_{t-2}] \end{aligned}$$

and estimated them by regression. For example, in the case of $k = 1$, μ can be estimated by regressing $\widetilde{h}_1(A_2), \cdots, \widetilde{h}_1(A_T)$ on $A_1, \cdots, A_{T-1}$. And we can solve the weights as

$$W = \mathbf{1} - H(H^T H)^{-1}[H^T \mathbf{1} - \mathbf{Df}]$$

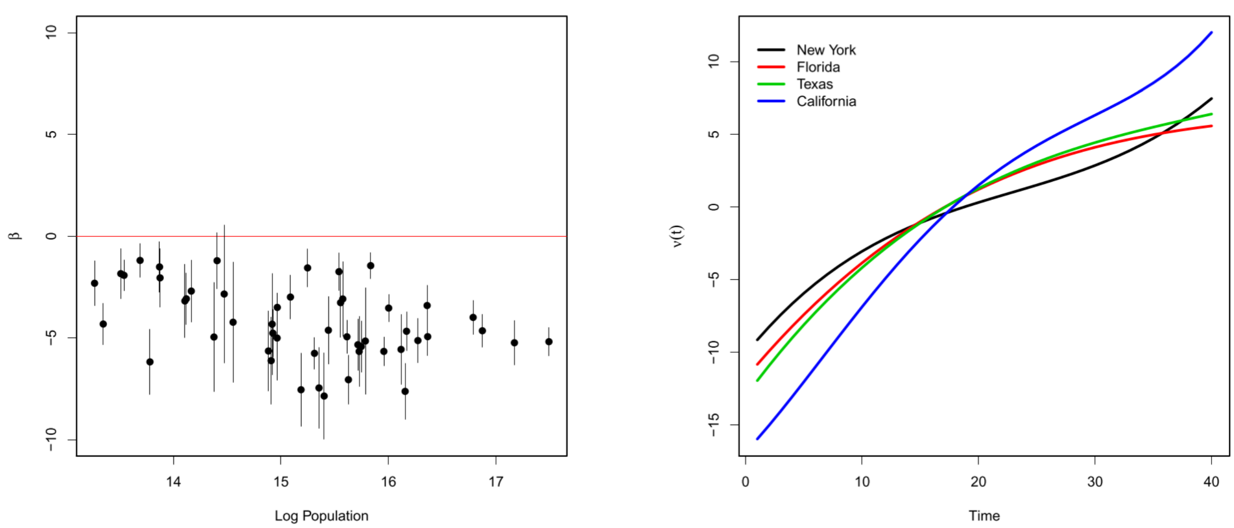
where $T = \sum_t W_t$, $W = (W_1, \cdots, W_T)$, $\mathbf{D} = (T, 0, \cdots, 0)^T$ and

$$H = \begin{bmatrix} 1 & H_{11} & \cdots & H_{1N} \\ 1 & H_{21} & \cdots & H_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & H_{T1} & \cdots & H_{TN} \end{bmatrix} \qquad \text{where} \quad H_{tj} = (\widetilde{h}_{1j}(A_t) - \widehat{\mu}_j)(\widetilde{h}_{2j}(Y_{t-1}) - \widehat{\nu}_j)$$

To sum it up, the algorithm is given by the following:

Algorithm:

1. Choose the order k of the Markov assumption.
2. Choose J pairs of functions $\{(\widehat{h}_{1j}(a), \widehat{h}_{1j}(y)) : j = 1, \cdots, J\}$.
3. Use regression to estimate $\mu_j = E[\widetilde{h}_{1j}(A_t)|A_{t-k}, \cdots, A_{t-1}]$ and $\nu_j = E[\widetilde{h}_{2j}(Y_{t-1})|A_{t-k-\delta-1}, \cdots, A_{t-\delta-1}, Y_{t-1-k}, \cdots, Y_{t-2}]$.
4. Compute the weights $W_1, \cdots, W_n$ by the last equation above.
5. Find the model $L_t = \beta \sum_{i=1}^{t-\delta} A_s + \nu(t) + \epsilon_t$ using weighted least squares with weigths $W_1, \cdots, W_n$.



The above graphs are directly taken out from the paper. The left one plots $\widehat{\beta}$ and 95% confidence interval verus the log population of each (55) state. These estimates are all negative (and they should be) since higher A_s implies less mobility. The right one plots $\widehat{\nu}(t)$ verus time in four popular states. The increase of death shown in this graph matches the common epidemic dynamics.

We can say that the algorithm performance intuitively makes sense, but we cannot say much about it without further sensitivity analysis.