Debugging Sandwich Variance DR ATE Estimator

--- 10월 5일 Version

〈What To Do〉

: Estimating equation "$\frac{1}{N}\sum_{i=1}^{N}\psi_i(\theta)$" = $\bar{\psi}(\theta)$ 라 할 때 각 Data 마다 (총 1000개의 set)

$\hat{\theta}$과 θ^* (true value) 에 대해 $(\bar{\psi}(\hat{\theta}),\ \bar{\psi}(\theta^*))$ 계산 ; 총 1000개의 pair에 대해

Sample Variance of $\bar{\psi}(\hat{\theta})$, Sample Variance of $\bar{\psi}(\theta^*)$ 근사한지 비교 (#of obs 바꾸어가며 확인!)

[Estimating equation of DR ATE] ⇒ Our Situation Version

$$\psi_i(\theta) = \begin{bmatrix} v_1 - \dfrac{\{E_i Y_i - (E_i - e_i)\,E[Y_i \mid E=1, X_i]\}}{e_i} \\ v_0 - \dfrac{\{(1-E_i)Y_i + (E_i - e_i)\,E[Y_i \mid E=0, X_i]\}}{1-e_i} \\ X_i(Y_i - \alpha^T X_i) \\ E_i \cdot X_i - \dfrac{e^{X_i^T\beta}}{1+e^{X_i^T\beta}} \cdot X_i \end{bmatrix}$$

(ψ_{v1}) (ψ_{v0}) (ψ_α) (ψ_β)

Covariance term

Outcome regression model의 model matrix

PS model의 model matrix

$\underset{\sim}{\theta} = (v_1, v_0, \underset{\sim}{\alpha}, \underset{\sim}{\beta})$

1) true parameter θ^* : $(E[Y^1], E[Y^0], \underset{\sim}{\alpha}, \underset{\sim}{\beta})$ 구하기 위해 Exposure 생성, Y 생성 process 확인

"$P(E=1 \mid X) = \text{logit}^{-1}(-2 + 0.01B + 0.01C)$, $Y = (\log 1.5, \log 2, \log 2)^T (B, C, E) + \varepsilon$" is our Situation / 따라서 $E[Y^1] = \log 1.5\, E[B] + \log 2\, E[C] + \log 2$, $E[Y^0] = \log 1.5\, E[B] + \log 2 \cdot E[C]$ 이고, 이때 Simulation 상황 상 $B \sim Ber(0.5)$, $C \sim N(0, 1^2)$ 이므로 $E[B] = 0.5$, $E[C] = 0$ 이다.

또한, $\underset{\sim}{\alpha}^{true} = (0, \log 1.5, \log 2, \log 2)^T$, $\underset{\sim}{\beta}^{true} = (-2, 0.01, 0.01)^T$ 이다.

$\Rightarrow E[Y|E=1, B, C] = \log 1.5 \times B + \log 2 \times C + \log 2 \cdot 1$, $E[Y|E=0, B, C] = \log 1.5 \times B + \log 2 \times C + \log 2 \cdot 0$

[Result]

1) # of obs = 100 일 때 : "z_hat"이 $\hat{z}(\hat{\theta})$, "z_true"가 $\hat{z}(\theta^*)$ 의미함.

```
> var(z_hat)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
 [1,]  4.389465e-33 -6.437336e-35  4.215076e-35  2.846144e-35 -2.038569e-34  8.883402e-35  1.008263e-27  3.574344e-28  2.397058e-28
 [2,] -6.437336e-35  2.056372e-33  1.023299e-34  8.011486e-35  6.256833e-35  7.809898e-35 -2.866743e-28  1.311884e-28  5.434004e-28
 [3,]  4.215076e-35  1.023299e-34  1.438671e-33  1.858350e-34  7.323514e-34  1.034111e-34 -5.221516e-28 -5.824666e-28  6.229053e-28
 [4,]  2.846144e-35  8.011486e-35  1.858350e-34  1.644123e-33  2.006840e-34  1.202562e-34 -7.017018e-28 -2.608980e-28  8.856073e-30
 [5,] -2.038569e-34  6.256833e-35  7.323514e-34  2.006840e-34  3.695359e-33  6.890613e-35 -6.722130e-29  2.252896e-28  7.835370e-28
 [6,]  8.883402e-35  7.809898e-35  1.034111e-34  1.202562e-34  6.890613e-35  2.187835e-33  7.500353e-28  5.556651e-28 -4.226186e-28
 [7,]  1.008263e-27 -2.866743e-28 -5.221516e-28 -7.017018e-28 -6.722130e-29  7.500353e-28  2.644563e-19  1.299015e-19 -1.291860e-20
 [8,]  3.574344e-28  1.311884e-28 -5.824666e-28 -2.608980e-28  2.252896e-28  5.556651e-28  1.299015e-19  1.083977e-19 -4.654881e-21
 [9,]  2.397058e-28  5.434004e-28  6.229053e-28  8.856073e-30  7.835370e-28 -4.226186e-28 -1.291860e-20 -4.654881e-21  1.032065e-19
> var(z_true)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
 [1,]  0.0833887638  5.913023e-03 -9.371410e-03 -9.292225e-03 -4.146079e-03 -1.619213e-03 -1.649143e-04 -2.621968e-04 -3.535159e-04
 [2,]  0.0059130229  1.650809e-02 -9.712081e-03 -4.641558e-05 -4.918403e-03 -4.410232e-04  9.334158e-05  5.516848e-05  6.057821e-05
 [3,] -0.0093714103 -9.712081e-03  9.518059e-03  1.104122e-03  4.831510e-03  3.089517e-04  1.504155e-05  1.380523e-05  1.937345e-05
 [4,] -0.0092922249 -4.641558e-05  1.104122e-03  1.107259e-03  4.993664e-04  1.706296e-04  2.885304e-05  3.478974e-05  4.616574e-05
 [5,] -0.0041460792 -4.918403e-03  4.831510e-03  4.993664e-04  4.868821e-03  6.703923e-05  7.789846e-05  2.296917e-05  3.742166e-05
 [6,] -0.0016192131 -4.410232e-04  3.089517e-04  1.706296e-04  6.703923e-05  9.285253e-03  6.772722e-06  3.498928e-06  1.094618e-04
 [7,] -0.0001649143  9.334158e-05  1.504155e-05  2.885304e-05  7.789846e-05  6.772722e-06  1.114492e-03  5.480218e-04  7.313380e-05
 [8,] -0.0002621968  5.516848e-05  1.380523e-05  3.478974e-05  2.296917e-05  3.498928e-06  5.480218e-04  5.404104e-04  2.706036e-05
 [9,] -0.0003535159  6.057821e-05  1.937345e-05  4.616574e-05  3.742166e-05  1.094618e-04  7.313380e-05  2.706036e-05  1.089614e-03
```

v_1 v_0 $\underset{\sim}{\alpha}$ $\underset{\sim}{\beta}$

$\Rightarrow$ 결과가 동일하지 않다.

2) # of obs = 1000 일 때 : "z_hat"이 $\hat{z}(\hat{\theta})$, "z_true"가 $\hat{z}(\theta^*)$ 의미

```
> var(z_hat)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
 [1,]  1.667902e-33 -6.295768e-35 -2.462950e-36  2.837396e-34  3.204476e-34  1.035774e-35  7.429677e-29  6.060258e-28  2.524582e-29
 [2,] -6.295768e-35  1.314118e-33 -5.414043e-35 -2.162372e-34  2.912507e-34 -5.544222e-35 -1.102571e-27 -5.105240e-28 -6.032019e-28
 [3,] -2.462950e-36 -5.414043e-35  1.457633e-33  2.358950e-34 -2.298436e-34 -1.162956e-35  8.912832e-28 -7.359467e-29 -8.450400e-28
 [4,]  2.837396e-34 -2.162372e-34  2.358950e-34  1.078603e-32 -7.616768e-34 -1.379121e-34 -1.190630e-28 -3.273858e-28 -1.687322e-27
 [5,]  3.204476e-34  2.912507e-34 -2.298436e-34 -7.616768e-34  5.212150e-32  1.410179e-35 -7.095134e-27 -5.234764e-27  1.465595e-27
 [6,]  1.035774e-35 -5.544222e-35 -1.162956e-35 -1.379121e-34  1.410179e-35  1.928611e-33 -5.825715e-28 -1.701258e-29 -3.044799e-29
 [7,]  7.429677e-29 -1.102571e-27  8.912832e-28 -1.190630e-28 -7.095134e-27 -5.825715e-28  3.212461e-19  1.360652e-19 -3.693205e-20
 [8,]  6.060258e-28 -5.105240e-28 -7.359467e-29 -3.273858e-28 -5.234764e-27 -1.701258e-29  1.360652e-19  1.084161e-19 -1.642348e-20
 [9,]  2.524582e-29 -6.032019e-28 -8.450400e-28 -1.687322e-27  1.465595e-27 -3.044799e-29 -3.693205e-20 -1.642348e-20  1.731612e-19
> var(z_true)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
 [1,]  9.028783e-03  3.564293e-04 -7.800870e-04 -1.007170e-03 -3.416734e-04  1.763024e-05  1.595356e-05 -3.203134e-06 -2.162356e-05
 [2,]  3.564293e-04  1.613930e-03 -9.224852e-04  2.219588e-05 -4.530020e-04  3.344007e-05  8.197218e-06 -8.742965e-07  3.675494e-06
 [3,] -7.800870e-04 -9.224852e-04  8.982600e-04  9.234995e-05  4.468994e-04 -4.708693e-05  8.094548e-06  6.578819e-06  8.343814e-06
 [4,] -1.007170e-03  2.219588e-05  9.234995e-05  1.195177e-04  4.196927e-05 -3.033916e-06  1.607829e-07  9.956277e-07  3.687121e-06
 [5,] -3.416734e-04 -4.530020e-04  4.468994e-04  4.196927e-05  4.880883e-04 -1.579659e-05  1.332457e-05  9.155539e-06  1.051984e-06
 [6,]  1.763024e-05  3.344007e-05 -4.708693e-05 -3.033916e-06 -1.579659e-05  1.003630e-03 -9.236314e-06 -5.516915e-06 -1.460192e-05
 [7,]  1.595356e-05  8.197218e-06  8.094548e-06  1.607829e-07  1.332457e-05 -9.236314e-06  9.705416e-05  5.075010e-05 -1.214042e-06
 [8,] -3.203134e-06 -8.742965e-07  6.578819e-06  9.956277e-07  9.155539e-06 -5.516915e-06  5.075010e-05  5.429635e-05 -2.965799e-06
 [9,] -2.162356e-05  3.675494e-06  8.343814e-06  3.687121e-06  1.051984e-06 -1.460192e-05 -1.214042e-06 -2.965799e-06  1.080202e-04
```

v_1 v_0 $\underset{\sim}{\alpha}$ $\underset{\sim}{\beta}$

3) # of obs = 10000 일때 : "z_hat" 이 $z(\hat{\theta})$, "z_true" 가 $z(\theta^*)$ 의미

```
> var(z_hat)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
[1,]  1.229843e-33  3.722968e-35 -1.407608e-35 -2.137418e-34  1.027190e-34  1.804522e-35  3.659307e-29  2.152663e-29 -1.185573e-29
[2,]  3.722968e-35  1.428129e-33  4.562452e-36  3.027295e-34 -1.003793e-33 -8.909173e-36 -7.853848e-29 -5.505058e-29  3.890480e-29
[3,] -1.407608e-35  4.562452e-36  1.354231e-33  5.696386e-34  7.365970e-34  4.678414e-35  1.213442e-28  3.043430e-29 -9.338959e-29
[4,] -2.137418e-34  3.027295e-34  5.696386e-34  1.313645e-31  9.850855e-33 -4.966956e-35 -4.439759e-28 -1.266826e-28  1.980347e-28
[5,]  1.027190e-34 -1.003793e-33  7.365970e-34  9.850855e-33  1.465646e-31 -3.305089e-34  9.418796e-28  5.223046e-28 -7.654207e-29
[6,]  1.804522e-35 -8.909173e-36  4.678414e-35 -4.966956e-35 -3.305089e-34  2.213848e-33  3.677862e-29 -2.970232e-29 -1.242169e-29
[7,]  3.659307e-29 -7.853848e-29  1.213442e-28 -4.439759e-28  9.418796e-28  3.677862e-29  1.820584e-21  8.548950e-22 -8.321334e-22
[8,]  2.152663e-29 -5.505058e-29  3.043430e-29 -1.266826e-28  5.223046e-28 -2.970232e-29  8.548950e-22  6.781305e-22 -3.641441e-22
[9,] -1.185573e-29  3.890480e-29 -9.338959e-29  1.980347e-28 -7.654207e-29 -1.242169e-29 -8.321334e-22 -3.641441e-22  1.313395e-21
> var(z_true)
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
[1,]  8.593965e-04  2.613297e-05 -8.057100e-05 -9.751055e-05 -4.037362e-05 -2.768898e-06  1.653826e-06  2.358612e-06 -3.565075e-06
[2,]  2.613297e-05  1.696687e-04 -1.011542e-04  3.021088e-06 -5.012064e-05 -4.524690e-06 -4.744274e-07  2.060799e-07 -6.464125e-07
[3,] -8.057100e-05 -1.011542e-04  9.808907e-05  9.572707e-06  4.917002e-05  2.284593e-06 -6.448335e-07 -7.432882e-07  1.212340e-06
[4,] -9.751055e-05  3.021088e-06  9.572707e-06  1.174723e-05  4.877054e-06  2.002866e-07 -2.985418e-07 -3.143512e-07  4.504490e-07
[5,] -4.037362e-05 -5.012064e-05  4.917002e-05  4.877054e-06  4.754964e-05 -7.940996e-07 -9.429928e-07 -5.190544e-07  1.368452e-06
[6,] -2.768898e-06 -4.524690e-06  2.284593e-06  2.002866e-07 -7.940996e-07  9.969840e-05 -7.358215e-07  3.328952e-07 -5.447920e-07
[7,]  1.653826e-06 -4.744274e-07 -6.448335e-07 -2.985418e-07 -9.429928e-07 -7.358215e-07  1.048796e-05  5.353965e-06 -1.193861e-07
[8,]  2.358612e-06  2.060799e-07 -7.432882e-07 -3.143512e-07 -5.190544e-07  3.328952e-07  5.353965e-06  5.240656e-06 -3.720360e-08
[9,] -3.565075e-06 -6.464125e-07  1.212340e-06  4.504490e-07  1.368452e-06 -5.447920e-07 -1.193861e-07 -3.720360e-08  1.084531e-05
```

v_1 v_0 $\underset{\sim}{\alpha}$ $\underset{\sim}{\beta}$

⇒ $\hat{\theta}$ 과 θ^* set 에 대해 계산한 "$\frac{1}{N}\sum_{i=1}^{N}\psi_i(\theta)$" 의 Sample Variance 가 다르다