Finding the Optimal Coefficient Forms

Caleb Leedy

13 February 2024

Summary

In this document, we do the following:

1. Vary the functional forms of the coefficients,
2. Check to ensure that our estimated variance is still accurate.

Idea

In `20240202-estcovmat.qmd`, we tested an estimated covariance matrix. The results suggest that estimating the covariance matrix suffers little penalty. In this report, we want to see if choosing different functional forms of f_1 and f_2 (that depend on the data) cause the covariance estimation to be biased.

Instead of using $f_1(X_1) = X_1$ and $f_2(X_2) = X_2$ as we did in `20240202-estcovmat.qmd` we want to use $f_1(X_1) = E[Y \mid X_1]$ and $f_2(X_2) = E[Y \mid X_2]$, *and* we need to also test them with estimated versions of these functions.

Simulation

Table 1: Results from simulations study with independent equally sized segments A_{11}, A_{10}, A_{01}, and A_{00} all of size $n = 250$. In this simulation we have the true mean of Y_2 equal to $\mu = 5$ and the covariance between e_1 and e_2 is $\rho = 0.5$. The goal is to estimate $E[Y_2] = \mu$. For the GLS estimation, we use the estimated covariance matrix $\hat{V}$ with f-functions $f_1 = X_1$, $f_2 = X_2$ and $f_3 = Y$. For the GLS optimal estimation we are trying $f_1 = E[Y \mid X_1]$, $f_2 = E[Y \mid X_2]$ and $f_3 = Y$. Each of these expectations are estimated with using linear regression. So $E[Y \mid X_i]$ is implemented as a regression estimator

Algorithm	Bias	SD	Tstat	Pval
Oracle	0.002	0.044	2.533	0.006
CC	0.003	0.090	1.350	0.089
IPW	0.003	0.090	1.350	0.089
GLS	0.003	0.052	2.281	0.011
GLSEstVar	0.003	0.052	2.286	0.011
GLSOpt	0.003	0.053	2.170	0.015

Conclusion

Is the covariance estimation still working? (By working we mean causing the estimator to be unbiased.)

- It looks like estimating the variance is not causing additional bias in the estimator. This is great.
- We may, however, need to check more complicated simulation setups because the simpicity of the current setup might make it look better than it actually is.
- One suprising thing is that the ``optimal'' estimator is not better than the other estimators. This could be because of the simplicity of the setup, but we may need to check it out.