The Problem with the Weighted Linear Model

Besides the semiparametric model, we have been comparing our results to an optimal weighted linear model. This model can be summarized as the following:

1. For each segment A_k with a distinct missing pattern, let $G_k(Z)$ be the observed data in segment A_k.
2. For each A_k, estimate $\theta = E[g(X, Y_1, Y_2)]$ as $\hat{\theta}_k = E[g \mid G_k(Z)]$.
3. Combine these estimates using the inverse variance weighted average of each $\hat{\theta}_k$:

$$\hat{\theta} = \sum_k w_k \hat{\theta}_k \text{ where } w_k = \frac{V(\hat{\theta}_k)}{\sum_k V(\hat{\theta}_k)}.$$

By linear model theory this is the BLUE in the linear model case, and this works really well in Simulation 1 in which

$$\begin{bmatrix} x_i \\ e_{1i} \\ e_{1i} \end{bmatrix} \overset{ind}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix} \right)$$
$$y_{1i} = x_i + e_{1i}$$
$$y_{2i} = \theta + x_i + e_{2i}$$
$$g(X, Y_1, Y_2) = Y_2$$

As demonstrated in Table 1, the WLS estimator outperforms all of the other estimators except the Oraclex estimator which uses all of the data and even some elements that are missing!

Table 1: True Theta is 5. $\text{Cov}(e_1, e_2) = 0.5$

Algorithm	Bias	SD	T-stat	P-val
Oracle	0.001	0.044	1.546	0.061
Oraclex	0.000	0.032	0.479	0.316
CC	0.002	0.058	1.549	0.061
IPW	0.006	0.210	1.504	0.066
WLS	0.001	0.040	1.207	0.114
Prop	0.002	0.051	1.911	0.028
OptSemi	0.002	0.051	1.964	0.025

However, in Simulation 2, in which we keep the same setup as Simulation 1, but change $g(X, Y_1, Y_2) = Y_1^2 Y_2$. In this case, we have

Table 2: True g is 10. $\text{Cov}(e_1, e_2) = 0.5$

Algorithm	Bias	SD	T-stat	P-val
Oracle	0.007	0.529	0.741	0.229
Oraclex	0.004	0.475	0.510	0.305
CC	0.023	0.824	1.518	0.065
IPW	0.031	0.915	1.846	0.032
WLS	-0.131	0.387	-18.504	0.000
Prop	0.011	0.635	0.912	0.181
SemiOpt	0.011	0.636	0.983	0.163

Table 2 clearly shows that the WLS estimator is biased, but since this is a weighted average of unbiased linear estimators, how can this be biased?

The answer comes from the fact that we weight based on the variance of each unbiased estimator. A closer examination of how the bias of one of the estimators from A_k affects variance can be seen in Figure 1.

Figure 1: A scatterplot of the Bias-Variance relationship for Section A_{11}, in which all of the data is observed. A similar relationship is found in each of the other sections too.

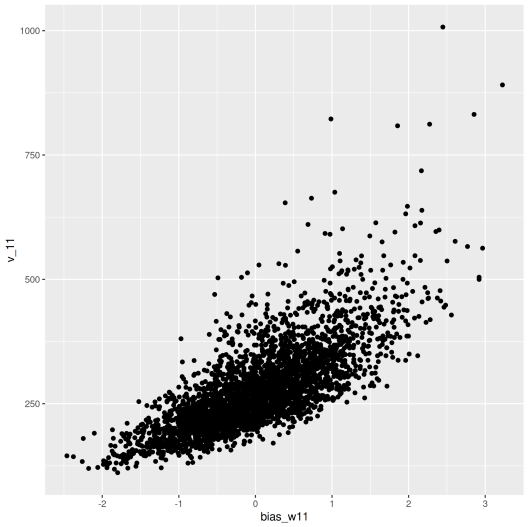

Figure 1 shows that the variance increases as the bias becomes more positive. This means that higher estimates will have a higher variance. Since we take the inverse variance weighted average of the four estimators, we are putting more weight on smaller estimators which will lead to a negative bias of our linear estimator which we observe in Table 2.