

# Optimal Estimation with Nuisance Parameters

Caleb Leedy

2024-02-16

## Summary

In this document, we do the following:

1. Estimate the optimal  $\theta$  with a known covariance function,
2. Estimate the optimal  $\theta$  with an estimated covariance function, and
3. Compare these results to other estimates.

## Simulation Setup

We consider a normal model for  $(X_1, X_2, Y)$  where  $X_1 \sim N(0, 1)$ , and  $e_1, e_2$  are from a bivariate normal distribution each with mean zero, variance one, and covariance  $\rho = 0.5$ . Then  $X_2 = X_1 + e_1$  and  $Y = \mu + X_1 + e_2$  where  $\mu = 5$ . Each observation is independent of the others. Instead of always observing all three variables, we observe them with missingness. The observed segments can be viewed in Table 1. Each segment is independent of each other and of the same size. We can view each segment as an independent simple random sample of the superpopulation of size  $n = 250$ .

Table 1: This table shows the segment names with each variable that they contain.

Segment	Observed Variables
$A_{11}$	$X_1, X_2, Y$
$A_{10}$	$X_1, X_2$
$A_{01}$	$X_1, Y$
$A_{00}$	$X_1$

## Estimation

The goal of the simulation is to estimate  $\theta = E[Y]$ . We also estimate several nuisance parameters  $\eta_1 = E[X_1]$  and  $\eta_2 = E[X_2]$ . All of these parameters can be estimated using GLS estimation.

### The Marginal GLS Model

The model that we propose is the following. Consider the model, for  $g = (\eta_1, \eta_2, \theta)'$

$$\hat{f} = Zg + e \text{ where}$$

$$f_1^{(11)} = n^{-1} \sum_{i=1}^n \frac{\delta_{11}}{\pi_{11}} g_1(x_{1i})$$

$$f_2^{(11)} = n^{-1} \sum_{i=1}^n \frac{\delta_{11}}{\pi_{11}} g_2(x_{2i})$$

$$f_3^{(11)} = n^{-1} \sum_{i=1}^n \frac{\delta_{11}}{\pi_{11}} g_3(y_i)$$

$$f_1^{(10)} = n^{-1} \sum_{i=1}^n \frac{\delta_{10}}{\pi_{10}} g_1(x_{1i})$$

$$f_2^{(10)} = n^{-1} \sum_{i=1}^n \frac{\delta_{10}}{\pi_{10}} g_2(x_{2i})$$

$$f_1^{(01)} = n^{-1} \sum_{i=1}^n \frac{\delta_{01}}{\pi_{01}} g_1(x_{1i})$$

$$f_3^{(01)} = n^{-1} \sum_{i=1}^n \frac{\delta_{01}}{\pi_{01}} g_2(y_i)$$

$$f_1^{(00)} = n^{-1} \sum_{i=1}^n \frac{\delta_{00}}{\pi_{00}} g_1(x_{1i})$$

and

$$\hat{f} = \begin{bmatrix} f_1^{(11)} \\ f_2^{(11)} \\ f_3^{(11)} \\ f_1^{(10)} \\ f_2^{(10)} \\ f_1^{(01)} \\ f_3^{(01)} \\ f_1^{(00)} \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, E[e] = 0, \text{ and } \text{Var}(e) = n^{-1} \begin{bmatrix} V_{11} & 0 & 0 & 0 \\ 0 & V_{10} & 0 & 0 \\ 0 & 0 & V_{01} & 0 \\ 0 & 0 & 0 & V_{00} \end{bmatrix}.$$

We also have

$$V_{11} = \begin{bmatrix} \frac{1}{\pi_{11}}E[g_1^2] - E[g_1]^2 & \frac{1}{\pi_{11}}E[g_1g_2] - E[g_1]E[g_2] & \frac{1}{\pi_{11}}E[g_1g_3] - E[g_1]E[g_3] \\ \frac{1}{\pi_{11}}E[g_1g_2] - E[g_1]E[g_2] & \frac{1}{\pi_{11}}E[g_2^2] - E[g_2]^2 & \frac{1}{\pi_{11}}E[g_2g_3] - E[g_2]E[g_3] \\ \frac{1}{\pi_{11}}E[g_1g_3] - E[g_1]E[g_3] & \frac{1}{\pi_{11}}E[g_2g_3] - E[g_2]E[g_3] & \frac{1}{\pi_{11}}E[g_3^2] - E[g_3]^2 \end{bmatrix},$$

$$V_{10} = \begin{bmatrix} \frac{1}{\pi_{10}}E[g_1^2] - E[g_1]^2 & \frac{1}{\pi_{10}}E[g_1g_2] - E[g_1]E[g_2] \\ \frac{1}{\pi_{10}}E[g_1g_2] - E[g_1]E[g_2] & \frac{1}{\pi_{10}}E[g_2^2] - E[g_2]^2 \end{bmatrix},$$

$$V_{01} = \begin{bmatrix} \frac{1}{\pi_{01}}E[g_1^2] - E[g_1]^2 & \frac{1}{\pi_{01}}E[g_1g_3] - E[g_1]E[g_3] \\ \frac{1}{\pi_{01}}E[g_1g_3] - E[g_1]E[g_3] & \frac{1}{\pi_{01}}E[g_3^2] - E[g_3]^2 \end{bmatrix}, \text{ and } V_{00} = \left[ \frac{1}{\pi_{00}}E[g_1^2] - E[g_1]^2 \right].$$

Using GLS directly, yields an estimator  $\hat{g} = (\hat{\eta}_1, \hat{\eta}_2, \hat{\theta})$ . Previously, we have used the estimator<sup>1</sup>

$$\hat{\theta}_M := [0, 0, 1] \cdot \hat{g}.$$

## The Conditional GLS Model

However, we can do better if we assume that the distribution of  $\hat{g}$  is a joint multivariate normal model and compute  $\hat{\theta} \mid \hat{\eta}_1, \hat{\eta}_2$ . We can represent this model<sup>2</sup> as  $\hat{\theta}_C := E[\hat{\theta} \mid \hat{\eta}]$ . In this representation, the initial goal is to minimize  $Q(\eta_1, \eta_2, \theta)$  where

---

<sup>1</sup>The ``M'' is short for marginal.

<sup>2</sup>The ``C'' is for conditional.

$$Q(\eta_1, \eta_2, \theta) = z'V^{-1}z \text{ for } z = \begin{bmatrix} f_1^{(11)} - \eta_1 \\ f_1^{(10)} - \eta_1 \\ f_1^{(01)} - \eta_1 \\ f_1^{(00)} - \eta_1 \\ f_2^{(11)} - \eta_2 \\ f_2^{(10)} - \eta_2 \\ f_2^{(01)} - \eta_2 \\ f_3^{(11)} - \theta \\ f_3^{(01)} - \theta \end{bmatrix} \text{ and } V = V(z).$$

More details about how we solve this model in in Appendix A. For each of these models we use the true covariance matrix and the estimated covariance matrix.

## Simulation Study

Table 2: 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 estimators with 'est', 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 estimators without we use the true covariance matrix. GLSM is the marginal model while GLSC is the conditional model.

Algorithm	Bias	SD	Tstat	Pval
Oracle	-0.002	0.044	-1.124	0.131
CC	-0.005	0.088	-1.820	0.035
IPW	-0.005	0.088	-1.820	0.035
GLSM	-0.002	0.053	-1.286	0.099
GLSMEst	-0.002	0.053	-1.271	0.102
GLSC	-0.002	0.053	-1.304	0.096
GLSCEst	-0.020	0.200	-3.135	0.001

## Results and Discussion

1. Estimating the covariance in our conditional model cause dependence issues.

2. This particular simulation does not differentiate between the GLS models. We might need a different setup to distinguish between the models (such as unequal segment size).

## Next Steps

1. Find the minimum variance estimator GLSC as a function of  $g$ .
2. Estimate the covariance matrix so that it does not interfere with estimation of  $\hat{f}$ .

Comment on the second bullet point. Currently, I estimate the covariance matrix by estimating the corresponding expected values separately using all of the available data. See Appendix B for more details. For example, to estimate  $E[X_1X_2]$  we use

$$\hat{E}[X_1X_2] = \frac{\sum_{i=1}^n (\delta_{10} + \delta_{11})(x_{1i}x_{2i})}{\sum_{i=1}^n \delta_{10} + \delta_{11}}.$$

I think that this may be overcome using REML. I will be working on this.

## Appendix A: Derivation of $\hat{\theta}_C$

The derivation in this appendix follows the idea in Zhou and Kim (2012). The initial goal is to minimize  $Q(\eta_1, \eta_2, \theta)$  where

$$Q(\eta_1, \eta_2, \theta) = z'V^{-1}z \text{ for } z = \begin{bmatrix} f_1^{(11)} - \eta_1 \\ f_1^{(10)} - \eta_1 \\ f_1^{(01)} - \eta_1 \\ f_1^{(00)} - \eta_1 \\ f_2^{(11)} - \eta_2 \\ f_2^{(10)} - \eta_2 \\ f_3^{(11)} - \theta \\ f_3^{(01)} - \theta \end{bmatrix} \text{ and } V = V(z).$$

We can notice that  $Q$  is the kernel of a multivariate normal distribution, which means that we can decompose it into a marginal and conditional part,

$$\min_{\eta, \theta} Q(\eta_1, \eta_2, \theta) = \min_{\eta, \theta} (Q(\eta) + Q(\theta | \eta)).$$

Since  $Q(\eta)$  is the kernel of a normal distribution we can apply the same technique to get

$$\min_{\eta, \theta} Q(\eta_1, \eta_2, \theta) = \min_{\eta, \theta} \{Q(\eta_1) + Q(\eta_2 | \eta_1) + Q(\theta | \eta_1, \eta_2)\}.$$

Then we can solve these sequentially.

1. We find  $\hat{\eta}_1 = \text{argmin} Q(\eta_1)$ .
2. We can substitute this back into  $Q(\eta_2 | \eta_1)$  to get the optimal  $\hat{\eta}_2$  of

$$\hat{\eta}_2 = \text{argmin} Q(\eta_2 | \hat{\eta}_1).$$

3. We can use the results of the previous two steps to get,

$$\hat{\theta} = \text{argmin} Q(\theta | \hat{\eta}_1, \hat{\eta}_2).$$

Unfortunately, the nonmonotone case does not have any nice simplifications like how the monotone case reduces to a regression estimator. This is due to the fact that the marginal and conditional distributions are generally multivariate normal distributions instead of a one-dimensional distribution. This causes the result to be weighted averages instead of difference estimators.

### Finding $\hat{\eta}_1$

We start with the fact that

$$\hat{\eta}_1 = \operatorname{argmin} Q(\eta_1) \text{ where } Q(\eta_1) = z'_{1:4} V_{1:4,1:4}^{-1} z_{1:4}.$$

This means that for the simulation study in this report, since each segment is independent,

$$Q(\eta_1) = \begin{bmatrix} \hat{f}_1^{(11)} - \eta_1 \\ \hat{f}_1^{(10)} - \eta_1 \\ \hat{f}_1^{(01)} - \eta_1 \\ \hat{f}_1^{(00)} - \eta_1 \end{bmatrix}' \begin{bmatrix} (\operatorname{Var}(\hat{f}_1^{(11)}))^{-1} & 0 & 0 & 0 \\ 0 & (\operatorname{Var}(\hat{f}_1^{(10)}))^{-1} & 0 & 0 \\ 0 & 0 & (\operatorname{Var}(\hat{f}_1^{(01)}))^{-1} & 0 \\ 0 & 0 & 0 & (\operatorname{Var}(\hat{f}_1^{(00)}))^{-1} \end{bmatrix} \begin{bmatrix} \hat{f}_1^{(11)} - \eta_1 \\ \hat{f}_1^{(10)} - \eta_1 \\ \hat{f}_1^{(01)} - \eta_1 \\ \hat{f}_1^{(00)} - \eta_1 \end{bmatrix}.$$

Since the dual of this problem is a GLS estimator with a diagonal variance matrix,  $\hat{\eta}_1$  is the inverse-variance weighted average of all of the individual estimators,

$$\hat{\eta}_1 = \frac{\frac{\hat{f}_1^{(11)}}{\operatorname{Var}(\hat{f}_1^{(11)})} + \frac{\hat{f}_1^{(10)}}{\operatorname{Var}(\hat{f}_1^{(10)})} + \frac{\hat{f}_1^{(01)}}{\operatorname{Var}(\hat{f}_1^{(01)})} + \frac{\hat{f}_1^{(00)}}{\operatorname{Var}(\hat{f}_1^{(00)})}}{\frac{1}{\operatorname{Var}(\hat{f}_1^{(11)})} + \frac{1}{\operatorname{Var}(\hat{f}_1^{(10)})} + \frac{1}{\operatorname{Var}(\hat{f}_1^{(01)})} + \frac{1}{\operatorname{Var}(\hat{f}_1^{(00)})}}.$$

### Finding $\hat{\eta}_2$

Similar to the previous step, we can notice that  $Q(\eta_2 \mid \eta_1)$  is the kernel of the conditional distribution

$$\begin{bmatrix} \hat{f}_2^{11} \\ \hat{f}_2^{10} \end{bmatrix} \sim N \left( \begin{bmatrix} \eta_2 \\ \eta_2 \end{bmatrix} + \Sigma_{\eta_2, \eta_1} \Sigma_{\eta_1}^{-1} \begin{bmatrix} \hat{f}_1^{(11)} - \eta_1 \\ \hat{f}_1^{(10)} - \eta_1 \\ \hat{f}_1^{(01)} - \eta_1 \\ \hat{f}_1^{(00)} - \eta_1 \end{bmatrix}, \Sigma_{\eta_2} - \Sigma_{\eta_2, \eta_1} \Sigma_{\eta_1}^{-1} \Sigma_{\eta_1, \eta_2} \right).$$

Since we have  $\Sigma_{\eta_1} = V_{1:4,1:4}$  and

$$\Sigma_{\eta_2, \eta_1} = \begin{bmatrix} \operatorname{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)}) & 0 & 0 & 0 \\ 0 & \operatorname{Cov}(\hat{f}_1^{(10)}, \hat{f}_2^{(10)}) & 0 & 0 \end{bmatrix},$$

we have

$$\Sigma_{\eta_2} - \Sigma_{\eta_2, \eta_1} \Sigma_{\eta_1}^{-1} \Sigma_{\eta_2, \eta_1} = \begin{bmatrix} \text{Var}(\hat{f}_2^{(11)}) - \frac{\text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)})^2}{\text{Var}(\hat{f}_1^{(11)})} & 0 \\ 0 & \text{Var}(\hat{f}_2^{(10)}) - \frac{\text{Cov}(\hat{f}_1^{(10)}, \hat{f}_2^{(10)})^2}{\text{Var}(\hat{f}_1^{(10)})} \end{bmatrix}.$$

This means that

$$\hat{\eta}_2 = \left( \begin{bmatrix} 1 & 1 \end{bmatrix} (\Sigma_{\eta_2} - \Sigma_{\eta_2, \eta_1} \Sigma_{\eta_1}^{-1} \Sigma_{\eta_2, \eta_1})^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} (\Sigma_{\eta_2} - \Sigma_{\eta_2, \eta_1} \Sigma_{\eta_1}^{-1} \Sigma_{\eta_2, \eta_1})^{-1} \begin{bmatrix} \hat{f}_2^{(11)} - \frac{\text{Cov}(\hat{f}_2^{(11)}, \hat{f}_1^{(11)})}{\text{Var}(\hat{f}_1^{(11)})} (\hat{f}_1^{(11)} - \hat{\eta}_1) \\ \hat{f}_2^{(10)} - \frac{\text{Cov}(\hat{f}_2^{(10)}, \hat{f}_1^{(10)})}{\text{Var}(\hat{f}_1^{(10)})} (\hat{f}_1^{(10)} - \hat{\eta}_1) \end{bmatrix}.$$

### Finding $\hat{\theta}$

Finally, to find  $\hat{\theta}$  we use a similar process as shown in the previous step. This time it is used to compute  $Q(\theta \mid \hat{\eta}_1, \hat{\eta}_2)$ . We can notice that  $Q(\theta \mid \hat{\eta})$  is the kernel of a conditional multivariate normal distribution,

$$\begin{bmatrix} \hat{f}_3^{(11)} \\ \hat{f}_3^{(01)} \end{bmatrix} \sim N \left( \begin{bmatrix} \theta \\ \theta \end{bmatrix} + \Sigma_{\theta, \eta} \Sigma_{\eta}^{-1} \begin{bmatrix} \hat{f}_1^{(11)} - \eta_1 \\ \hat{f}_1^{(10)} - \eta_1 \\ \hat{f}_1^{(01)} - \eta_1 \\ \hat{f}_1^{(00)} - \eta_1 \\ \hat{f}_2^{(11)} - \eta_2 \\ \hat{f}_2^{(10)} - \eta_2 \end{bmatrix}, \Sigma_{\theta} - \Sigma_{\theta, \eta} \Sigma_{\eta}^{-1} \Sigma_{\eta, \theta} \right).$$

In this case we have,

$$\Sigma_{\theta} = \begin{bmatrix} V(\hat{f}_3^{(11)}) & 0 \\ 0 & V(\hat{f}_3^{(10)}) \end{bmatrix}, \Sigma_{\theta, \eta} = \begin{bmatrix} \text{Cov}(\hat{f}_3^{(11)}, \hat{f}_1^{(11)}) & 0 & 0 & 0 & \text{Cov}(\hat{f}_3^{(11)}, \hat{f}_2^{(10)}) & 0 \\ 0 & 0 & \text{Cov}(\hat{f}_3^{(01)}, \hat{f}_1^{(01)}) & 0 & 0 & 0 \end{bmatrix},$$

and,



$$\Sigma_\eta = \begin{bmatrix} V(\hat{f}_1^{(11)}) & 0 & 0 & 0 & \text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)}) & 0 \\ 0 & V(\hat{f}_1^{(10)}) & 0 & 0 & 0 & \text{Cov}(\hat{f}_1^{(10)}, \hat{f}_2^{(10)}) \\ 0 & 0 & V(\hat{f}_1^{(01)}) & 0 & 0 & 0 \\ 0 & 0 & 0 & V(\hat{f}_1^{(00)}) & 0 & 0 \\ \text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)}) & 0 & 0 & 0 & V(\hat{f}_2^{(11)}) & 0 \\ 0 & \text{Cov}(\hat{f}_1^{(10)}, \hat{f}_2^{(10)}) & 0 & 0 & 0 & V(\hat{f}_2^{(10)}) \end{bmatrix}.$$

Hence,

$$\hat{\theta} = \left( \begin{bmatrix} 1 & 1 \end{bmatrix} (\Sigma_\theta - \Sigma_{\theta,\eta} \Sigma_\eta^{-1} \Sigma_{\eta,\theta})^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} (\Sigma_\theta - \Sigma_{\theta,\eta} \Sigma_\eta^{-1} \Sigma_{\eta,\theta})^{-1} \left( \begin{bmatrix} \hat{f}_3^{(11)} \\ \hat{f}_3^{(01)} \end{bmatrix} - \Sigma_{\theta,\eta} \Sigma_\eta^{-1} \begin{bmatrix} \hat{f}_1^{(11)} - \hat{\eta}_1 \\ \hat{f}_1^{(10)} - \hat{\eta}_1 \\ \hat{f}_1^{(01)} - \hat{\eta}_1 \\ \hat{f}_1^{(00)} - \hat{\eta}_1 \\ \hat{f}_2^{(11)} - \hat{\eta}_2 \\ \hat{f}_2^{(10)} - \hat{\eta}_2 \end{bmatrix} \right).$$

While this might still be slightly unclear, I have been able to write down the closed form solution to this estimator, which is given below in Figure 1 as a function of  $\eta_1$ . (Including  $\hat{\eta}_1$  made it too big for the page.)

$$\frac{-y^1 - (x_1^1 - \eta_1) \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{-(c(x_1^1 x_2^1))^2 + v(x_1^1) v(x_2^1)} + \frac{c(x_1^1 y^1) v(x_2^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} \right) + x_2^1 \left( \frac{\frac{c(x_1^2 x_2^2)(x_1^2 - \eta_1) - x_2^2}{v(x_1^2)} + \frac{c(x_1^1 x_2^1)(x_1^1 - \eta_1) - x_2^1}{v(x_1^1)}}{\frac{(c(x_1^2 x_2^2))^2 - v(x_1^2)}{v(x_1^2)} - \frac{(c(x_1^1 x_2^1))^2 - v(x_1^1)}{v(x_1^1)}} \right) \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} - \frac{c(x_1^1 y^1) v(x_2^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} \right)}{\frac{-c(x_1^1 y^1) \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{-(c(x_1^1 x_2^1))^2 + v(x_1^1) v(x_2^1)} + \frac{c(x_1^1 y^1) v(x_2^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} \right) + c(x_2^1 y^1 \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} - \frac{c(x_2^1 y^1) v(x_1^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} \right) - v(y^1)} + \frac{c(x_1^2 y^2)(x_1^2 - \eta_1) - y^2}{v(x_1^2)} - \frac{(c(x_1^2 y^2))^2 - v(y^2)}{v(x_1^2)}}{\frac{c(x_1^1 y^1) \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{-(c(x_1^1 x_2^1))^2 + v(x_1^1) v(x_2^1)} - \frac{c(x_1^1 y^1) v(x_2^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} \right) + c(x_2^1 y^1 \left( \frac{c(x_1^1 x_2^1) c(x_1^1 y^1)}{(c(x_1^1 x_2^1))^2 - v(x_1^1) v(x_2^1)} + \frac{c(x_2^1 y^1) v(x_1^1)}{-(c(x_1^1 x_2^1))^2 + v(x_1^1) v(x_2^1)} \right) - v(y^1) + \frac{(c(x_1^2 y^2))^2 - v(y^2)}{v(x_1^2)}}}$$

Figure 1: A closed form solution to  $\hat{\theta}$ .

## Appendix B: Computing the Conditional Covariance Matrix

For the simulation study in this report the covariance matrix for the conditional model  $\hat{\theta}_C$ . Since each segment is considered to be independent, using the current estimator  $\hat{\theta}_C$  we only need to compute the variance of each  $\hat{f}_j$  and the following covariances:  $\text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)})$ ,  $\text{Cov}(\hat{f}_1^{(11)}, \hat{f}_3^{(11)})$ ,  $\text{Cov}(\hat{f}_2^{(11)}, \hat{f}_3^{(11)})$ ,  $\text{Cov}(\hat{f}_1^{(10)}, \hat{f}_2^{(10)})$ , and  $\text{Cov}(\hat{f}_1^{(01)}, \hat{f}_3^{(01)})$ .

The variances of each  $\hat{f}_j$  are relatively straightforward. As an example, I will derive  $\text{Var}(\hat{f}_1^{(11)})$ .

$$\text{Var}(\hat{f}_1^{(11)}) = n^{-1} \pi_{11}^{-1} (E[X_1^2] - E[X_1]^2).$$

For other values of  $\hat{f}_j$  we simply use the correct  $\pi_{ij}$  and variable from  $Z = (X_1, X_2, Y)$ . In a similar manner, the covariance  $\text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)})$ , can be computed as

$$\text{Cov}(\hat{f}_1^{(11)}, \hat{f}_2^{(11)}) = n^{-1} \pi_{11}^{-1} (E[X_1 X_2] - E[X_1] E[X_2]).$$

For the other covariances, we can substitute the corresponding  $\pi_{ij}$  and other variables from  $Z$ .

We using the true covariance, I use the population covariance that I know from the construction of the problem. To estimate the covariance matrix, I estimate each required expectation (such as  $E[X_1 X_2]$ ,  $E[X_1]$ , and  $E[X_2]$  for the previous covariance) using all of the available data. This means that

$$\hat{E}[X_1 X_2] = \frac{\sum_{i=1}^n (\delta_{11} + \delta_{10}) x_{1i} x_{2i}}{\sum_{i=1}^n (\delta_{11} + \delta_{10})}, \hat{E}[X_1] = n^{-1} \sum_{i=1}^n x_{1i} \text{ and } \hat{E}[X_2] = \frac{\sum_{i=1}^n (\delta_{01} + \delta_{11}) x_{2i}}{\sum_{i=1}^n \delta_{01} + \delta_{11}}.$$

## References

Zhou, Ming, and Jae Kwang Kim. 2012. "An Efficient Method of Estimation for Longitudinal Surveys with Monotone Missing Data." *Biometrika* 99 (3): 631--48.