Optimal Estimation with Nuisance Parameters

Caleb Leedy

2024-02-16

Summary

In this document, we do the following:

1. Estimate the optimal θ with a known covariance function,
2. Estimate the optimal θ 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 \$ = \$ 0.5. Then $X_2 = X_1 + e_1$ and $Y = \mu + X_1 + e_1$ where \$ = \$ 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 model that we propose is the following. Consider the model, for $g = (\eta_1, \eta_2, \theta)'$

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

$$
\begin{aligned}
f_1^{(11)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{11}}{\pi_{11}} g_1(x_{1i}) \\
f_2^{(11)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{11}}{\pi_{11}} g_2(x_{2i}) \\
f_3^{(11)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{11}}{\pi_{11}} g_3(y_i) \\
f_1^{(10)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{10}}{\pi_{10}} g_1(x_{1i}) \\
f_2^{(10)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{10}}{\pi_{10}} g_2(x_{2i}) \\
f_1^{(01)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{01}}{\pi_{01}} g_1(x_{1i}) \\
f_3^{(01)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{01}}{\pi_{01}} g_2(y_i) \\
f_1^{(00)} &= n^{-1} \sum_{i=1}^{n} \frac{\delta_{00}}{\pi_{00}} g_1(x_{1i})
\end{aligned}
$$

and

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

We also have

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

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

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

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

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

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[2] as $\hat{\theta}_C := E[\hat{\theta} \mid \hat{\eta}]$. For each of these models we use the true covariance matrix and the estimated covariance matrix. For a full discussion of the derivation of $\hat{\theta}_C$ see Appendix A.

[1]The ``M'' is short for marginal.

[2]The ``C'' is for conditional.

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 $\widehat{V}$ 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	-1.268	0.103
GLSCEst	-0.013	0.055	-7.397	0.000

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 $\hat{f}$.

Appendix A: Derivation of $\widehat{\theta}_C$

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,

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

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

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

Then we can solve these sequentially.

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

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

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

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

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 $\widehat{\eta}_1$

We start with the fact that

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

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

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

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

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

Finding $\widehat{\eta}_2$

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

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

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

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

we have

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

This means that

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

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,

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

and,

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

Hence,

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

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 η_1. (Including $\hat{\eta}_1$ made it too big for the page.)

$$
-\frac{\dfrac{-Y^1-(X_1^1-\eta_1)\left(\dfrac{C(X_1^1X_2^1)C(X_2^1Y^1)}{-\left(C(X_1^1X_2^1)\right)^2+V(X_1^1)V(X_2^1)}+\dfrac{C(X_1^1Y^1)V(X_2^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}\right)+\left(X_2^1-\dfrac{\dfrac{\frac{C(X_1^2X_2^2)(X_1^2-\eta_1)}{V(X_1^2)}-X_2^2}{\frac{\left(C(X_1^2X_2^2)\right)^2}{V(X_1^2)}-V(X_2^2)}+\dfrac{\frac{C(X_1^1X_2^1)(X_1^1-\eta_1)}{V(X_1^1)}-X_2^1}{\frac{\left(C(X_1^1X_2^1)\right)^2}{V(X_1^1)}-V(X_2^1)}}{-\dfrac{1}{\frac{\left(C(X_1^2X_2^2)\right)^2}{V(X_1^2)}-V(X_2^2)}-\dfrac{1}{\frac{\left(C(X_1^1X_2^1)\right)^2}{V(X_1^1)}-V(X_2^1)}}\right)\left(\dfrac{C(X_1^1X_2^1)C(X_1^1Y^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}-\dfrac{C(X_2^1Y^1)V(X_1^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}\right)}{-C(X_1^1Y^1)\left(\dfrac{C(X_1^1X_2^1)C(X_2^1Y^1)}{-\left(C(X_1^1X_2^1)\right)^2+V(X_1^1)V(X_2^1)}+\dfrac{C(X_1^1Y^1)V(X_2^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}\right)+C(X_2^1Y^1)\left(\dfrac{C(X_1^1X_2^1)C(X_1^1Y^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}-\dfrac{C(X_2^1Y^1)V(X_1^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}\right)-V(Y^1)}+\dfrac{\frac{C(X_1^3Y^2)(X_1^3-\eta_1)}{V(X_1^3)}-Y^2}{\frac{\left(C(X_1^3Y^2)\right)^2}{V(X_1^3)}-V(Y^2)}}{\dfrac{1}{C(X_1^1Y^1)\left(-\dfrac{C(X_1^1X_2^1)C(X_2^1Y^1)}{-\left(C(X_1^1X_2^1)\right)^2+V(X_1^1)V(X_2^1)}-\dfrac{C(X_1^1Y^1)V(X_2^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}\right)+C(X_2^1Y^1)\left(\dfrac{C(X_1^1X_2^1)C(X_1^1Y^1)}{\left(C(X_1^1X_2^1)\right)^2-V(X_1^1)V(X_2^1)}+\dfrac{C(X_2^1Y^1)V(X_1^1)}{-\left(C(X_1^1X_2^1)\right)^2+V(X_1^1)V(X_2^1)}\right)-V(Y^1)}+\dfrac{1}{\frac{\left(C(X_1^3Y^2)\right)^2}{V(X_1^3)}-V(Y^2)}}
$$

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

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.