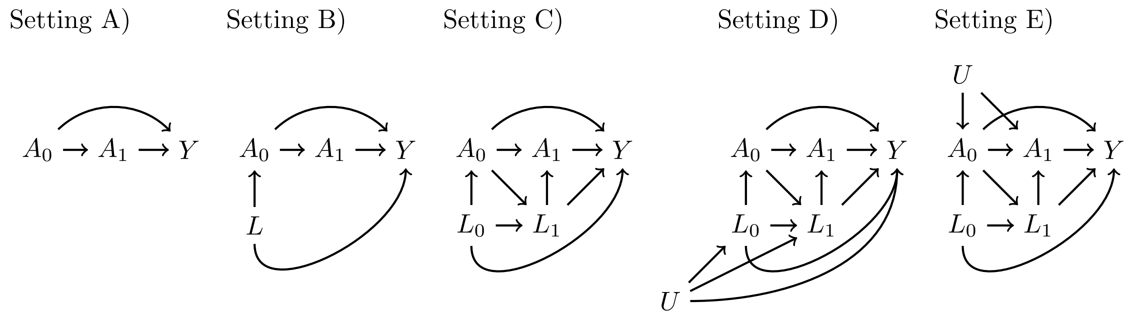
Info 6751. Fall 2022. Problem Set 8. Due on Canvas by 5pm on 24 Oct.

Submit a PDF. For Part 2, either embed your code in the PDF or include as an attachment.

1 Causal identification in dynamic settings (30 points)

In each DAG below, the researcher observes the outcome Y and the variables $\{\vec{L}_t, A_t\}$ for every time period t. The researcher does not observe U.



The researcher wants to know the expected potential outcome $\text{E}(Y^{a_0,a_1})$ under an intervention to set $A_0 = a_0$ and $A_1 = a_1$. In which of settings A–E can this causal estimand be identified using

1.1. (10 points) The simple mean $\text{E}(Y \mid A_0 = a_0, A_1 = a_1)$

1.2. (10 points) Inverse probability weights applied in a marginal structural model, adjusting at each period for the observed past

Now focus on Setting D.

1.3. (10 points) For Setting D, explain how L_1 complicates the quest to jointly identify the causal effects of A_0 and A_1 using our usual adjustment strategies.

2 Causal estimation in a dynamic setting (20 points)

This problem is an extension of the class exercise from Tuesday. The goal is for you to practice estimating a marginal structural model in a dynamic setting designed to be as simple as possible.

Suppose a teacher observes students who are struggling academically ($L_t = 1$) or not ($L_t = 0$) in every time period t. The teacher can assign students to receive extra support ($A_t = 1$) or not ($A_t = 0$). The outcome Y is a subsequent measure of skills (e.g., can the student read a picture book to the teacher). The attached `pset8.csv` contains a simulated setting with $n = 128$ students. It is not identical to the one from class, but as in class:

- Treatment $A_t = 1$ is assigned with higher probabilities to students who are struggling $L_t = 0$.
- Students who are treated in one period $A_0 = 1$ are never struggling in the next period $L_1 = 0$.

Assume the causal structure represented by Setting D above. Using these data,

2.1. (3 points) Fit a logistic regression for each A_t, including as predictors all pre-treatment variables entered without interactions.

2.2. (3 points) Predict the probability of $A_t = 1$ at each time period for every unit. Remember that in R you will need the `type = "response"` option. For grading, what is the predicted value of A_0 for the first unit in the sample?

2.3. (3 points) Define the generalized propensity score π_{it} at each time period t for each unit i. When $A_{it} = 1$ this will be the number from 2.2. When $A_{it} = 0$, it will be $1 - ($`that number`$)$.

2.4. (3 points) Define a weight for each unit: $w_i = \frac{1}{\pi_{i0}} \frac{1}{\pi_{i1}}$. For grading, what is the weight on the first unit in the sample?

2.5. (3 points) Estimate a marginal structural model of the form $\text{E}(Y^{a_0,a_1}) = \alpha + \beta_0 a_0 + \beta_1 a_1$. Report the coefficients.

2.6. (5 points) What does your model estimate for the average causal effect $\text{E}(Y^{1,1} - Y^{0,0})$?