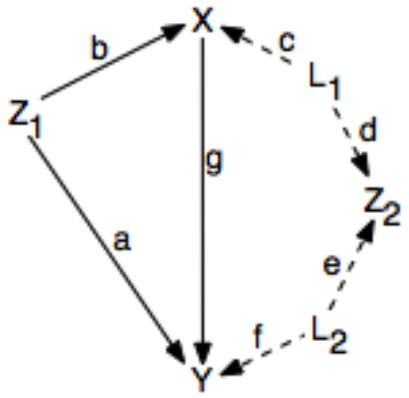
Ralph's "kitchen sink" fallacy

Ralph is pondering identification of the causal effect of X on Y for the causal structure below. The path coefficients indicate a linear structure ($Z_1 = U_{z_1}, X = bZ_1 + cL_1 + U_x, Z_2 = dL_1 + eL_2 + U_{z_2}, Y = aZ_1 + fL_2 + gX + U_y$).



Ralph knows unblocked back-door paths (into X connecting to Y) confound identification of the causal effect of interest (an omitted, correlated variable problem). Further, there is a temptation to include all available (measured) variables (the "kitchen sink" approach) as covariates in the analysis. However, this strategy may confound identification of the causal effect.

Suggested:

1. Which covariates should be included to identify the causal effect of X on Y? Explain.

2. If Z_1 is included as a covariate, determine the estimand (regression coefficient) $r_{YX \cdot Z_1}$.

3. If Z_1 and Z_2 are included as covariates, determine the estimand (regression coefficient) $r_{YX \cdot Z_1 Z_2}$. Suppose $g = 0$.

4. Based on conditional independencies, explain whether this causal structure can be identified from data. (hint: colliders block paths — create independencies — when excluded from the conditioning set but open paths — create dependencies — when included in the conditioning set).

5. Suppose $g = 0$. Repeat 4.

6. Suppose $Z_0 \xrightarrow{h} X$ is added to the above graph (Z_0 is an instrumental variable). Repeat 4. What issues with 4 does this help remedy? Does an instrumental variable strategy where Z_1 and Z_2 are not measured identify the causal effect of X on Y?