

# A Causal Zoo

2023-09-24

## Contents

1	A Causal Zoo	2
2	DAG - not in lower triangular form	3
3	DAG - in lower triangular form	4
4	Variance matrix	4
5	Some models to try	5
6	‘Fitting’ the models	6
7	Which model is best?	10
8	What’s happening with IVs?	12
	References	13

## List of Figures

1	A DAG . . . . .	2
2	$Var(\hat{\beta}_x)$ and $E(\hat{\beta}_x)$ : variance versus bias. . . . .	8
3	Variance versus bias with DAG . . . . .	9
4	Adjusted measure of fit versus bias. Which model(s) would you choose? . . . . .	11
5	A simple DAG with an IV . . . . .	12

# 1 A Causal Zoo

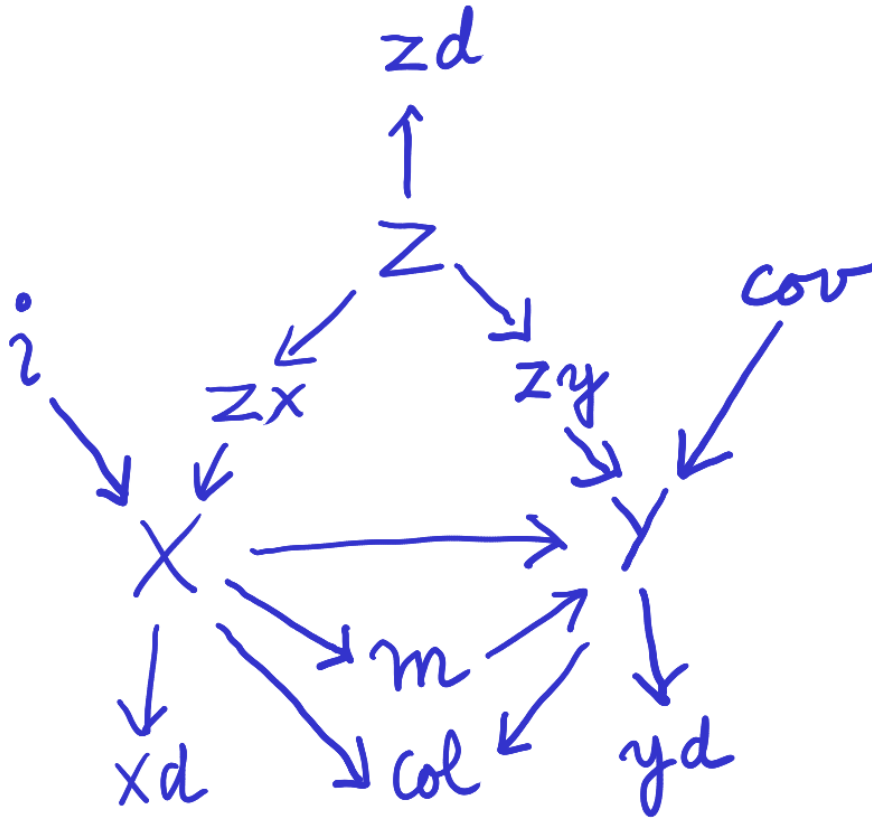


Figure 1: A DAG

```
nams <- c('z', 'zx', 'zy', 'cov', 'x', 'y',  
          'm', 'i', 'xd', 'yd', 'zd',  
          'col')  
mat <- matrix(0, length(nams), length(nams))  
rownames(mat) <- nams  
colnames(mat) <- nams  
  
# confounding back-door path  
  
mat['zx', 'z'] <- 3  
mat['zy', 'z'] <- 3  
mat['x', 'zx'] <- 1  
mat['y', 'zy'] <- 2  
  
# direct effect of x on y  
  
mat['y', 'x'] <- 3  
  
# indirect effect: Note that the causal effect is 3 + 1 x 1 = 4  
  
mat['m', 'x'] <- 1  
mat['y', 'm'] <- 1
```

```

# Instrumental variable
mat['x','i'] <- 1

# 'Covariate'
mat['y','cov'] <- 2

# descendant of X
mat['xd','x'] <- 1

# descendant of Y
mat['yd','y'] <- 1

# descendant of z -- imperfect control
mat['zd','z'] <- 2

# collider

mat['col','y'] <- 1
mat['col','x'] <- 1

# independent SD of error for every variable
diag(mat) <- 1

# but make SD of Y different
mat['y','y'] <- 4
mat['i','i'] <- 2.4

```

## 2 DAG - not in lower triangular form

```

mat
      z zx zy cov x y m  i xd yd zd col
z    1  0  0   0 0 0 0 0.0 0 0 0  0
zx   3  1  0   0 0 0 0 0.0 0 0 0  0
zy   3  0  1   0 0 0 0 0.0 0 0 0  0
cov  0  0  0   1 0 0 0 0.0 0 0 0  0
x    0  1  0   0 1 0 0 1.0 0 0 0  0
y    0  0  2   2 3 4 1 0.0 0 0 0  0
m    0  0  0   0 1 0 1 0.0 0 0 0  0
i    0  0  0   0 0 0 0 2.4 0 0 0  0
xd   0  0  0   0 1 0 0 0.0 1 0 0  0
yd   0  0  0   0 0 1 0 0.0 0 1 0  0
zd   2  0  0   0 0 0 0 0.0 0 0 1  0
col  0  0  0   0 1 1 0 0.0 0 0 0  1

```

### 3 DAG - in lower triangular form

Expressing the DAG in lower triangular form makes it easy to iteratively work out the variance matrix.

```
dag <- permute_to_dag(mat) # can be permuted to lower-diagonal form
dag
```

```
      i z zx x m cov zy y col zd yd xd
i    2.4 0 0 0 0 0 0 0 0 0 0 0
z    0.0 1 0 0 0 0 0 0 0 0 0 0
zx   0.0 3 1 0 0 0 0 0 0 0 0 0
x    1.0 0 1 1 0 0 0 0 0 0 0 0
m    0.0 0 0 1 1 0 0 0 0 0 0 0
cov  0.0 0 0 0 0 1 0 0 0 0 0 0
zy   0.0 3 0 0 0 0 1 0 0 0 0 0
y    0.0 0 0 3 1 2 2 4 0 0 0 0
col  0.0 0 0 1 0 0 0 1 1 0 0 0
zd   0.0 2 0 0 0 0 0 0 0 0 1 0
yd   0.0 0 0 0 0 0 0 0 1 0 0 1
xd   0.0 0 0 1 0 0 0 0 0 0 0 1
attr(,"class")
[1] "dag"      "matrix" "array"
```

### 4 Variance matrix

```
covld(dag)
```

```
      i z zx x m cov zy y col zd yd xd
i    5.76 0 0 5.76 5.76 0 0 23.04 28.8 0 23.04 5.76
z    0.00 1 3 3.00 3.00 0 3 18.00 21.0 2 18.00 3.00
zx   0.00 3 10 10.00 10.00 0 9 58.00 68.0 6 58.00 10.00
x    5.76 3 10 16.76 16.76 0 9 85.04 101.8 6 85.04 16.76
m    5.76 3 10 16.76 17.76 0 9 86.04 102.8 6 86.04 16.76
cov  0.00 0 0 0.00 0.00 1 0 2.00 2.0 0 2.00 0.00
zy   0.00 3 9 9.00 9.00 0 10 56.00 65.0 6 56.00 9.00
y    23.04 18 58 85.04 86.04 2 56 473.16 558.2 36 473.16 85.04
col  28.80 21 68 101.80 102.80 2 65 558.20 661.0 42 558.20 101.80
zd   0.00 2 6 6.00 6.00 0 6 36.00 42.0 5 36.00 6.00
yd   23.04 18 58 85.04 86.04 2 56 473.16 558.2 36 474.16 85.04
xd   5.76 3 10 16.76 16.76 0 9 85.04 101.8 6 85.04 17.76
```

## 5 Some models to try

```
fmlas <- list(  
  y ~ x,                # with confounding  
  y ~ x + z,           # unconfounded  
  y ~ x + zy,          # unconfounded using generating model  
  y ~ x + zx,          # unconfounded using assignment model  
  y ~ x + zx + zy,     # 'doubly robust'  
  y ~ x + zy + cov,    # adding a covariate unrelated to x  
  y ~ x + z + m,       # adding a mediator  
  y ~ x + z + xd,      # adding a descendant of X  
  y ~ x + z + yd,      # adding a descendant of Y  
  y ~ x + z + yd + cov, # adding a descendant of Y and a covariate  
  y ~ x + z + col,     # adding a collider  
  y ~ x + z + i,       # adding an instrumental variable  
  y ~ x + z + i + cov, # adding an instrumental variable and a covariate  
  y ~ x + xd,          # adding a descendant of x  
  y ~ x + i,           # using an instrumental variable as a control  
  y ~ x + zd,          # imperfect control for confounding  
  y ~ x + zd + cov,    # imperfect control + covariate  
  y ~ x + zd + xd,     # imperfect control + descendant of x  
  y ~ x + zd + i,      # bias amplification  
  y ~ x | i            # instrumental variable using two-stage least squares  
)
```

## 6 'Fitting' the models

```
fmlas %>%
  lapply(coefx, dag) %>%
  lapply(as.data.frame) %>%
  do.call(rbind.data.frame, .) -> df

df <- within(
  df,
  { # label positions for plotting
    pos <- ifelse(grepl('IV|zy|m|zd.*z', label), 2, 4)
    pos2 <- ifelse(grepl('yd$|z . xd|zd$', label), 2, pos)
    pos2 <- ifelse(grepl('yd', label), 3, pos2)
    pos2 <- ifelse(grepl('yd.*cov$', label), 1, pos2)
  }
)
pdf <- df
sapply(pdf, is.numeric) %>%
  {pdf[,.] <-< round(pdf[,.], 3)}
pdf[, c(6,1,4,2,3,5)] %>% print(row.names=F)
```

	label	beta_x	sd_factor	sd_e	sd_x_avp	var_e_adj
	y ~ x	5.074	1.577	6.455	4.094	43.156
	y ~ x + z	4.000	1.795	5.000	2.786	26.852
	y ~ x + zy	4.000	1.557	4.583	2.943	22.556
	y ~ x + zx	4.000	2.057	5.348	2.600	30.719
	y ~ x + zx + zy	4.000	1.763	4.583	2.600	23.423
	y ~ x + zy + cov	4.000	1.401	4.123	2.943	18.962
	y ~ x + z + m	3.000	5.205	4.899	0.941	26.769
	y ~ x + z + xd	4.000	5.312	5.000	0.941	27.885
	y ~ x + z + yd	0.154	0.846	0.981	1.159	1.072
y ~ x + z + yd + cov	0.182	0.904	0.977	1.081	1.107	
	y ~ x + z + col	-0.808	1.024	0.981	0.958	1.072
	y ~ x + z + i	4.000	3.536	5.000	1.414	27.885
y ~ x + z + i + cov	4.000	3.240	4.583	1.414	24.360	
	y ~ x + xd	5.074	6.645	6.455	0.971	44.755
	y ~ x + i	5.636	1.693	5.617	3.317	33.882
	y ~ x + zd	4.377	1.796	5.554	3.092	33.129
y ~ x + zd + cov	4.377	1.676	5.181	3.092	29.942	
	y ~ x + zd + xd	4.377	5.837	5.554	0.951	34.403
	y ~ x + zd + i	4.947	2.752	5.366	1.949	32.111
	y ~ x (IV = i)	4.000	3.254	7.810	2.400	63.179

```
cap = "$Var(\hat{\beta}_x)$ and $E(\hat{\beta}_x)$: variance versus bias."
```

```
df %>%
```

```
  xyplot(log2(sd_factor^2) ~ beta_x, . , font = 2, cex = 1,  
         scales = list(y = list(at=seq(-3,10), labels=2^(5+seq(-3,10)))),  
         xlim = c(-1.6, 6.5),  
         pch = 20,  
         xlab = TeX('$E(\hat{\beta}_x)$'),  
         ylab = TeX('$Var(\hat{\beta}_x)$ using relative $n$ to get equivalent power'),  
         labs = sub('y ~ ', '', .$label),  
         pos = .$pos) +  
  layer(panel.grid(h=-1,v=-1)) +  
  layer(panel.abline(v = 4, lty = 3)) +  
  layer(panel.text(..., labels = labs, pos = pos))
```







## 7 Which model is best?

**It depends on the purpose of the analysis!** Thanks to Hugh McCague for the idea of including the following figure to illustrate how focusing on predictive power does not lead to a suitable model to estimate the causal effect of X.

```
caption <- "Adjusted measure of fit versus bias. Which model(s) would you choose?"
```

```
df %>%
  subset(!grepl('i2', label)) %>%
  xyplot(var_e_adj ~ beta_x, . , font = 2,
         #scales = list(y = list(at=seq(-3,10), labels=2^(5+seq(-3,10)))),
         # xlim = c(-.6, 6),
         pch = 16,
         xlab = TeX('$E(\hat{\beta})$'),
         ylab = TeX('Goodness of fit using an equivalent to adjusted R-squared (smaller is better)'),
         labs = sub('y ~ ', '', .$label),
         pos = .$pos2) +
  layer(panel.text(..., labels = labs, pos = pos)) +
  layer(panel.grid(h=-1,v=-1)) +
  layer(panel.abline(v = 4, lty = 3))
```

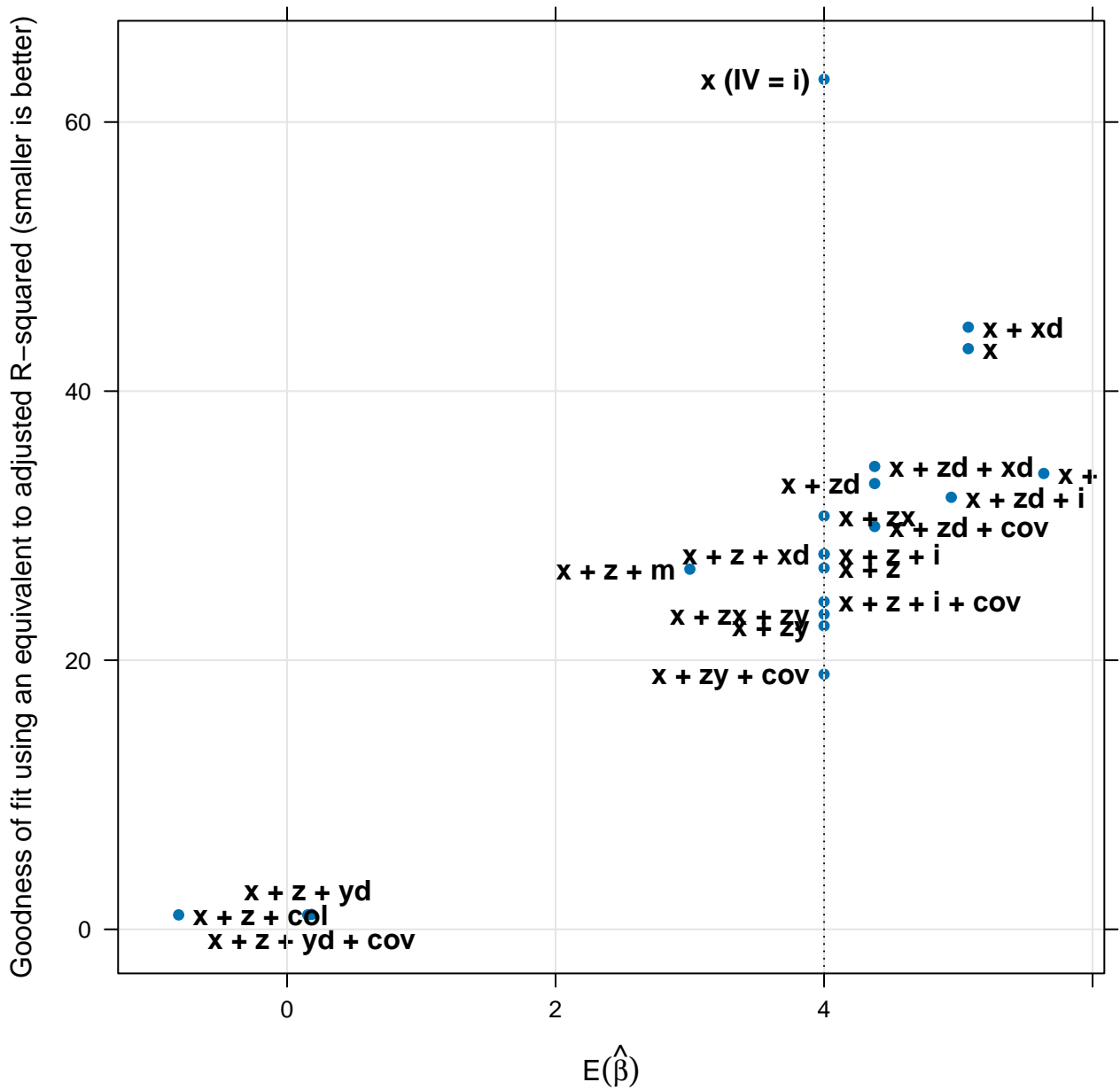


Figure 4: Adjusted measure of fit versus bias. Which model(s) would you choose?

## 8 What's happening with IVs?

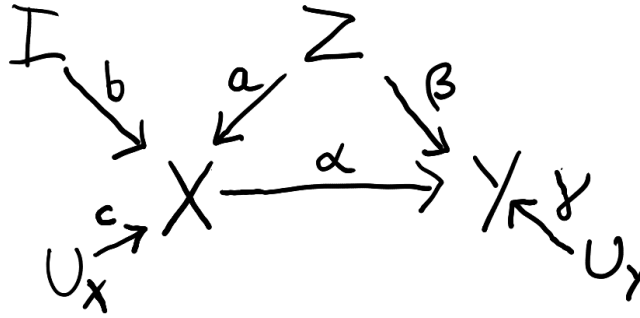


Figure 5: A simple DAG with an IV

Let's assume multivariate normality and build a variance matrix for  $Z, I, X, Y$ .

We can scale  $I, Z$  and  $X$  so they have unit variance and zero means. This eliminates irrelevant nuisance parameters.

Since  $I$  is an instrument for the confounding effect of  $Z$ :

$$\text{Var} \begin{pmatrix} Z \\ I \\ X \end{pmatrix} = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ a & b & 1 \end{pmatrix}$$

with  $a^2 + b^2 \leq 1$ .

Focus first on the *assignment model*, i.e. the model that determines the value of  $X$  from the values of  $Z, I$  and  $U_X$ .

Letting  $c^2 = 1 - a^2 - b^2$ ,  $c^2$  represent the portion of the variance in  $X$  that is not attributed to the instrument,  $I$ , nor to the confounder,  $Z$ , define

$$\rho_I = \frac{b^2}{b^2 + c^2}$$

the proportion of the variance in  $X$  not due to  $Z$  that is 'explained' by  $I$ .

For an instrument that captures all of the variation not due to the confounder,  $c^2 = 0$  and  $\rho_I = 1$ .

Focusing next on the model generating  $Y$ , let

$$Y = \alpha X + \beta Z + \gamma \varepsilon$$

with  $\varepsilon \sim N(0, 1)$ , independent of other variables.

The variance matrix is:

$$\text{Var} \begin{pmatrix} Z \\ I \\ X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 & a & a\alpha + \beta \\ 0 & 1 & b & b\alpha \\ a & b & 1 & \alpha + a\beta \\ a\alpha + \beta & b\alpha & \alpha + a\beta & v_{yy} \end{pmatrix}$$

where  $v_{yy} = \alpha^2 + \beta^2 + 2a\alpha\beta + \sigma_\varepsilon^2$

We can verify that the regression coefficients for the regression of  $Y$  on  $X$  and  $Z$  are

$$\begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha + a\beta \\ a\alpha + \beta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

The variance of the least-squares estimator of  $\alpha$  based on a regression on  $X$  and the confounder  $Z$  is:

$$\begin{aligned}\text{Var}(\hat{\alpha}) &\approx \frac{1}{n} \frac{\sigma_{\epsilon}^2}{1 - a^2} \\ &= \frac{1}{n} \frac{\gamma^2}{b^2 + c^2}\end{aligned}$$

The asymptotic expectation of the instrumental variable estimator  $\tilde{\alpha}$  is

$$\sigma_{IX}^{-1} \sigma_{IY} = \frac{1}{b} \times b\alpha = \alpha$$

The variance of  $\tilde{\alpha}$  is (Fox 2016, 241):

$$\text{Var}(\tilde{\alpha}) \approx \frac{1}{n} \sigma_{\epsilon IV}^2 \sigma_{IX}^{-1} \sigma_{II} \sigma_{XI}^{-1} = \frac{1}{n} (\beta^2 + \gamma^2) \frac{1}{b^2}$$

Thus, the variance inflation factor – which is the same as the ‘sample size inflation factor to achieve the same power’ – using IV estimation instead of controlling for a confounder (assuming that both approaches are available) is:

$$\begin{aligned}IVVIF &= \frac{\text{Var}(\tilde{\alpha})}{\text{Var}(\hat{\alpha})} \\ &= \frac{\beta^2 + \gamma^2}{b^2} / \frac{\gamma^2}{b^2 + c^2} \\ &= \frac{\beta^2 + \gamma^2}{\gamma^2} / \frac{b^2}{b^2 + c^2} \\ &= 1 / \left( \frac{\gamma^2}{\gamma^2 + \beta^2} \times \frac{b^2}{b^2 + c^2} \right) \\ &= \left( 1 + \frac{\beta^2}{\gamma^2} \right) \times \left( 1 + \frac{c^2}{b^2} \right) \\ &= \frac{1}{1 - R_{Y,Z|X}^2} \times \frac{1}{R_{X,I|Z}^2}\end{aligned}$$

The first term is structural in the sense that it is a consequence of the problem, specifically the degree of confounding relative to the residual error variance in the model. For a given problem, the IV has no impact on this, so it represents a lower bound for the IVVIF. The second term clarifies that it is not the *correlation of the IV with X* directly that affects the IVVIF, but its **partial correlation** adjusted for the relationship of X with confounders.

In conclusion: In any situation where you have a choice, controlling for confounders will do better than using a corresponding IV, i.e. an IV that annihilates the confounder. Fitting with IVs does not take the same advantage of a model with a small error variance in the same way that a regression model does.

The lower bound for error variance created by the confounder could swamp the benefit of small residual variance in the generating model. In contrast, a regression model takes full proportional advantage of a reduction in residual error.

## References

Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Sage Publications.