simulations_confouding

Michael Cork

1/20/2022

Count data

Confounding

We now look to generate the exposure data using different specifications of the relationship between confounding and exposure based on work by Xiao. We generate E using six different specifications that rely on the function $\gamma(\mathbf{C}) = -0.8+$ $(0.1, 0.1, -0.1, 0.2, 0.1, 0.1)$ C. Specifically, 1) $W = 9 \times \gamma(\mathbf{C}) + 17 + N(0,5)$; 2) $W = 15 \times \gamma(\mathbf{C}) + 22 + T(2)$; 3) $W = 9 \times \gamma(\mathbf{C}) + 3/2C_3^2 + 15 + N(0,5)$ 4) $W = 49 \times \frac{\exp(\gamma(\mathbf{C}))}{1+\exp(\gamma(\mathbf{C}))} - 6 + N(0,5)$; 22 5) $W = 42 \times \frac{1}{1+\exp(\gamma(\mathbf{C})\})} - 18 + N(0,5)$; 6) $W = 7 \times \log(\gamma(\mathbf{C})) + 13 + N(0,4)$;

One thing to note is that for 6), in actuality it is using $W = 7 \times \log(|\gamma(\mathbf{C})|) + 13 + N(0,4)$. I then ran through three different iterations of the sample size (N = 200, 1000, 5000) and then fit the following outcome model:

$$
\begin{aligned}
Y|E,C &\sim Pois(\mu(E,C)) \\
log(\mu(E,C)) &= 2 + 0.1 * E - (0.2, 0.2, 0.3, -0.1, 0.2, 0.2) * C
\end{aligned}
$$

gps_mod	sample_size	model	bias	mse
1	200	linear_model	2.733619e-01	1.992652e+00
1	200	gam_model	2.359080e-02	4.262503e+00
1	1000	linear_model	4.400800e-03	5.115241e-01
1	1000	gam_model	1.038210e-02	1.061761e+00
2	200	linear_model	1.021310e-01	4.367799e+01
2	200	gam_model	9.081375e-01	1.051411e+02
2	1000	linear_model	-6.659149e-01	4.484542e+01
2	1000	gam_model	-2.246836e-01	7.662117e+01
3	200	linear_model	2.161620e-02	2.323701e-01
3	200	gam_model	1.039950e-01	1.510932e+00
3	1000	linear_model	1.871913e-01	4.872672e+00
3	1000	gam_model	1.550040e-01	9.240206e+00
4	200	linear_model	-1.119050e-02	7.605035e-01
4	200	gam_model	-2.094760e-02	1.519948e+00
4	1000	linear_model	-8.931430e-02	6.154136e-01
4	1000	gam_model	-1.733474e-01	1.296533e+00
5	200	linear_model	-1.640017e-01	1.177963e+00
5	200	gam_model	1.460340e-02	7.029202e+00
5	1000	linear_model	2.937790e-02	2.662051e-01
5	1000	gam_model	-1.319360e-02	6.212679e-01
6	200	linear_model	3.202174e-01	4.202221e+00
6	200	gam_model	9.662872e-01	5.947636e+01
6	1000	linear_model	1.478495e+00	1.775422e+03

gps_mod	sample_size	model	bias	mse
6	1000	gam_model	6.405137e+06	7.939458e+15

Questions: What other types of outcome model should I use to make sure nonlinear confounding is occurring?

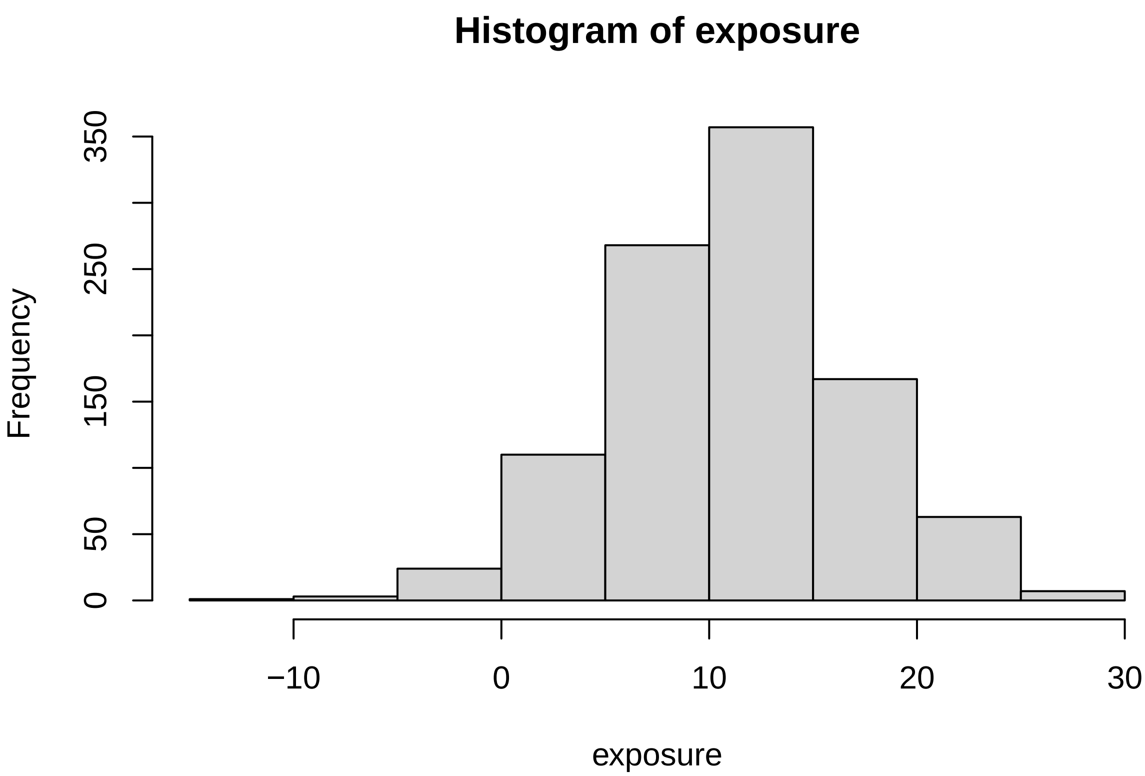
Types of confounding

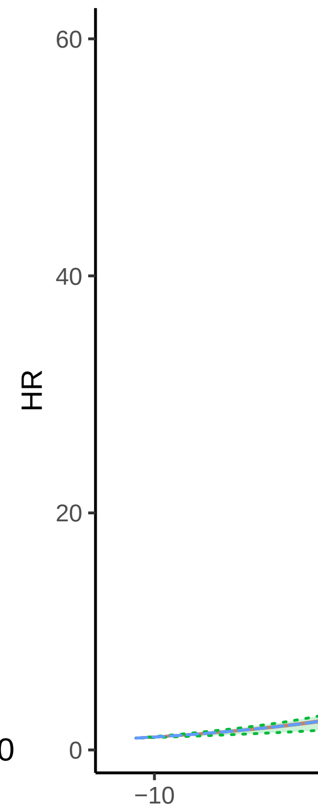
Count data with sublinear relationship

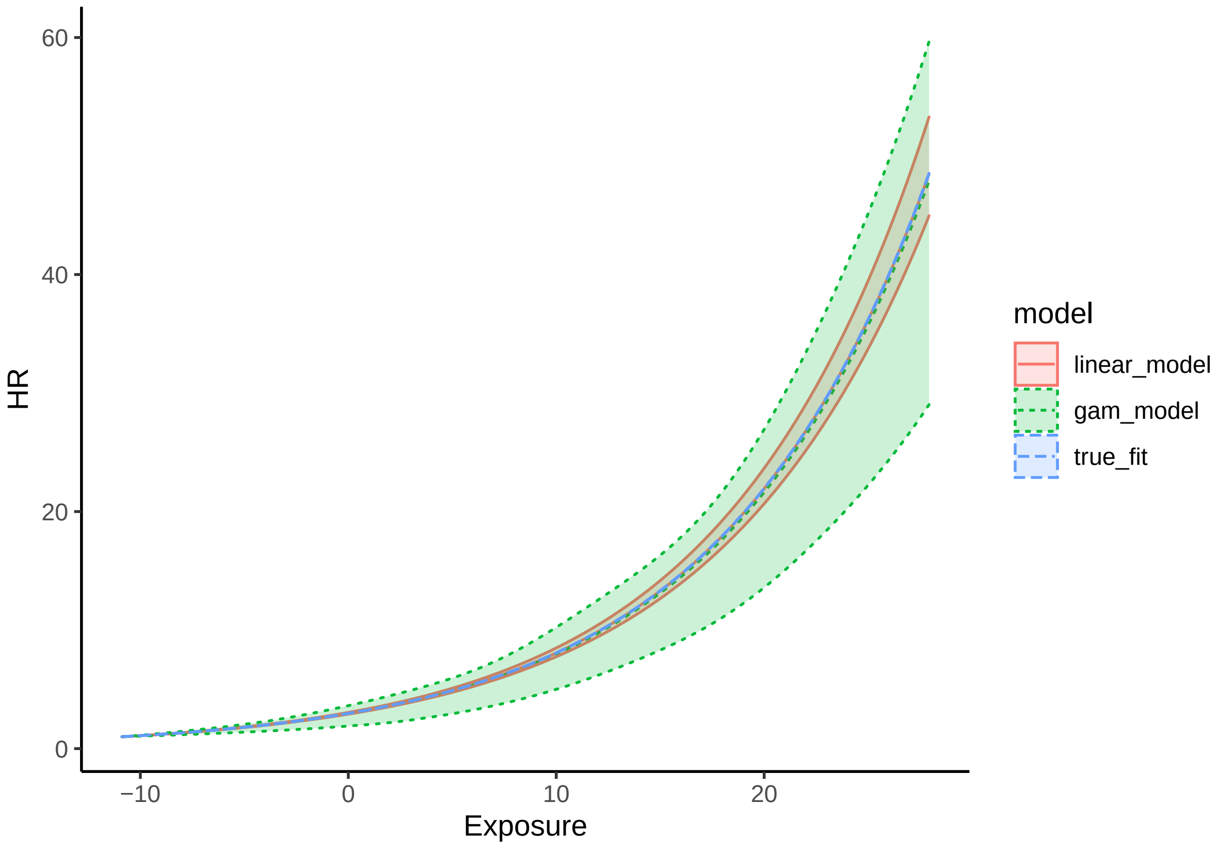
See equation above

```
##       5%
## 1.524179
```

```
##     95%
## 20.7137
```






60
40
20
0
HR
−10
0
10
20
Exposure
model
linear_model
gam_model
true_fit