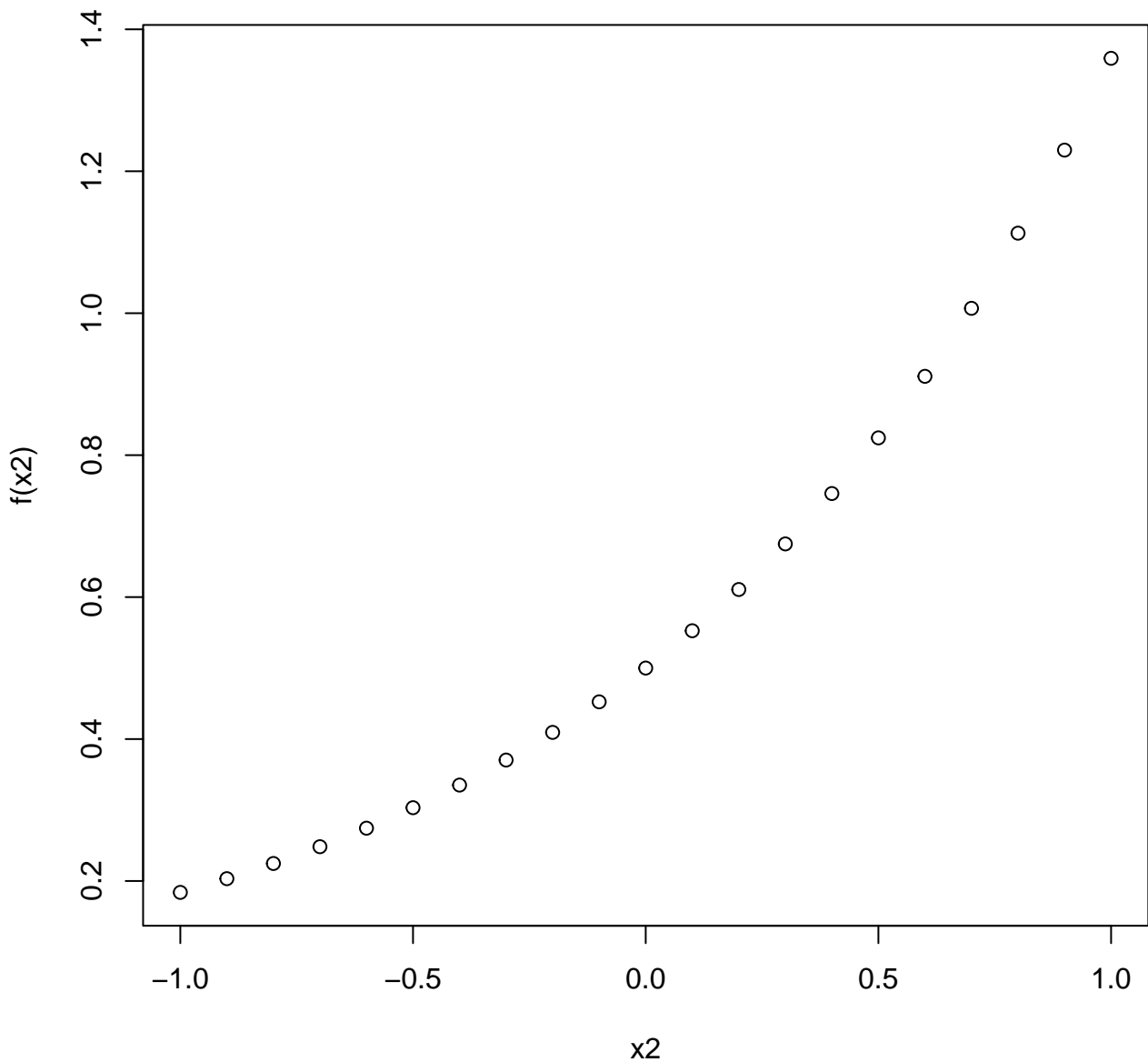
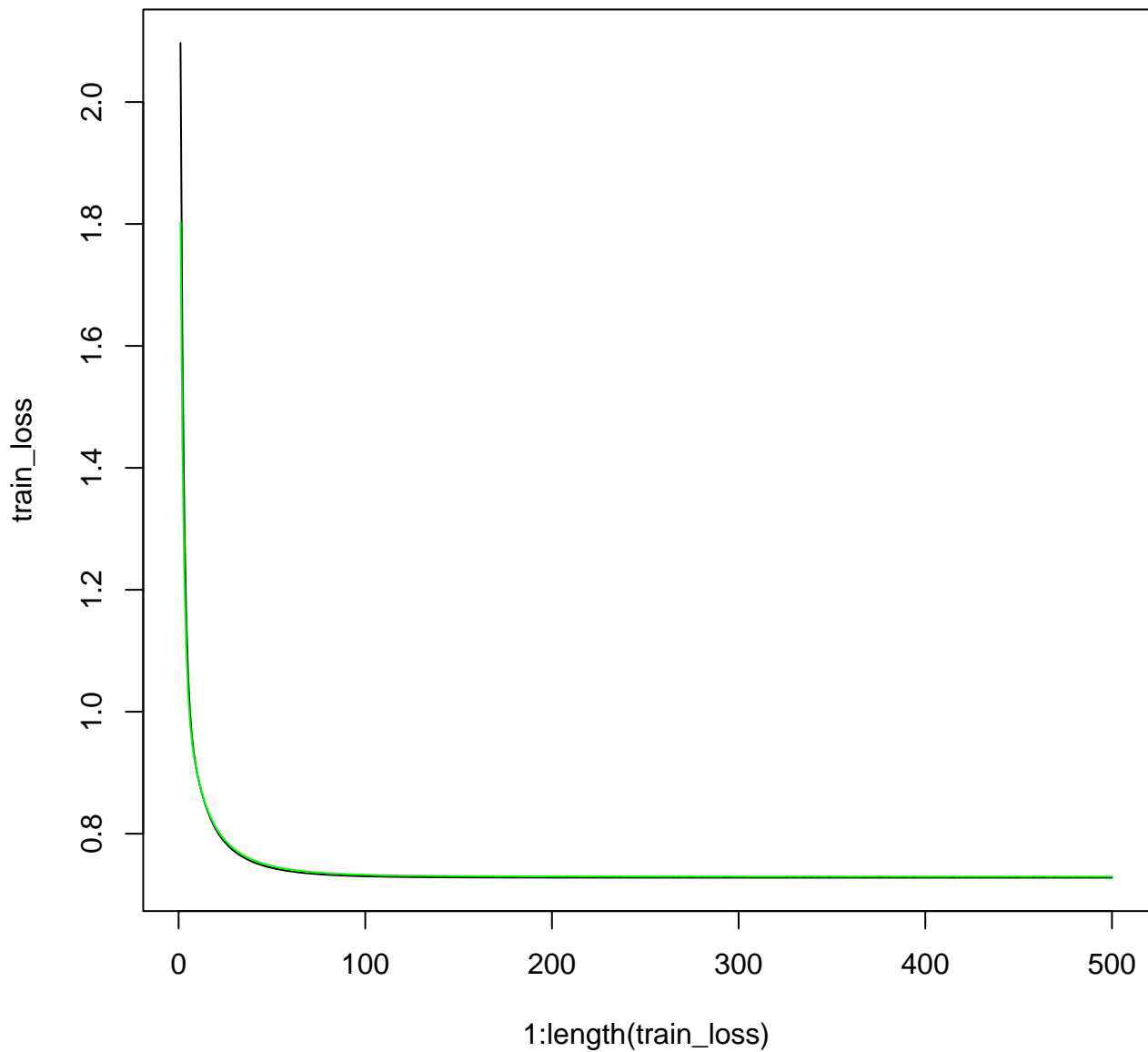


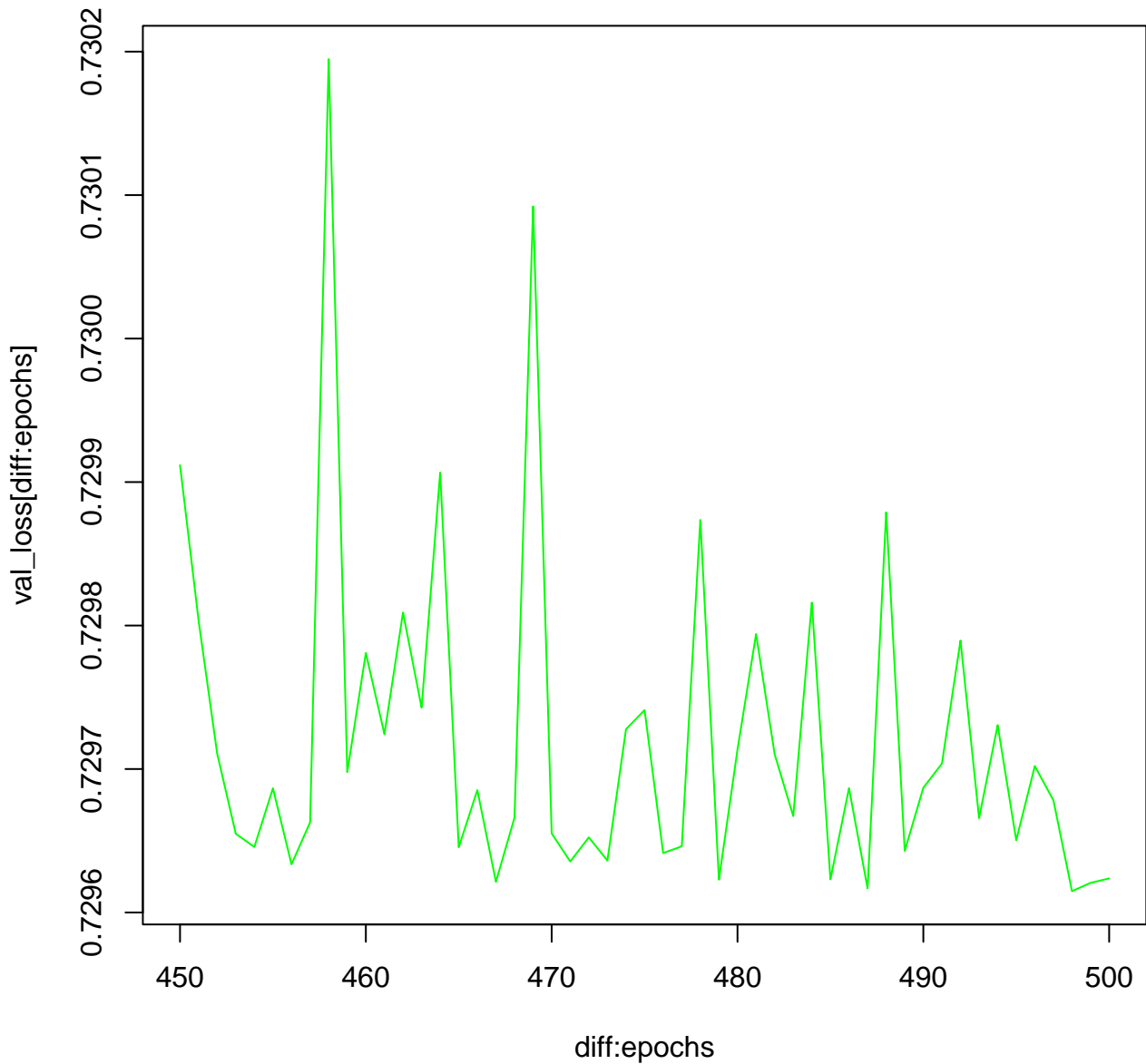
DGP influence of x2 on x3

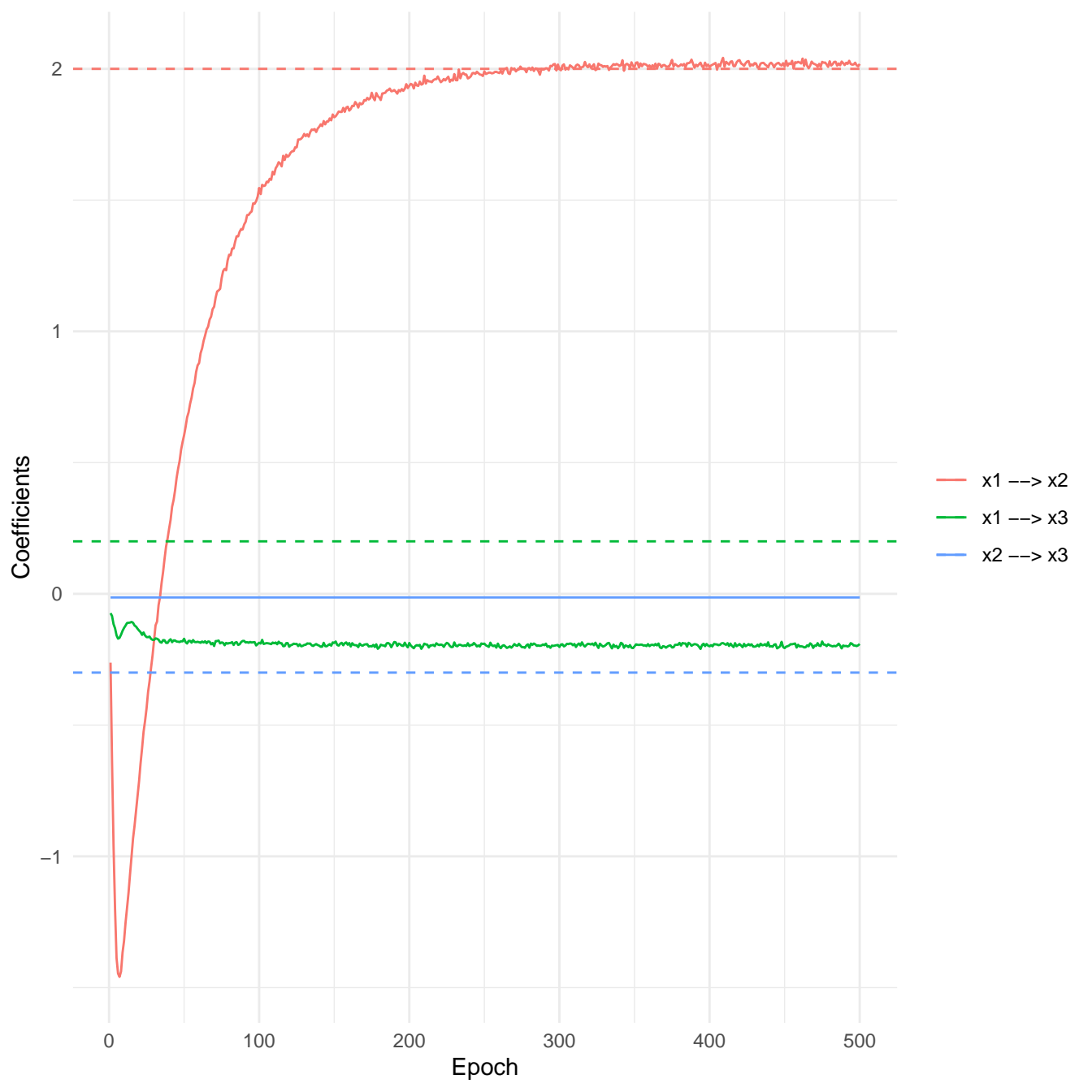


Normal Training (green is valid)

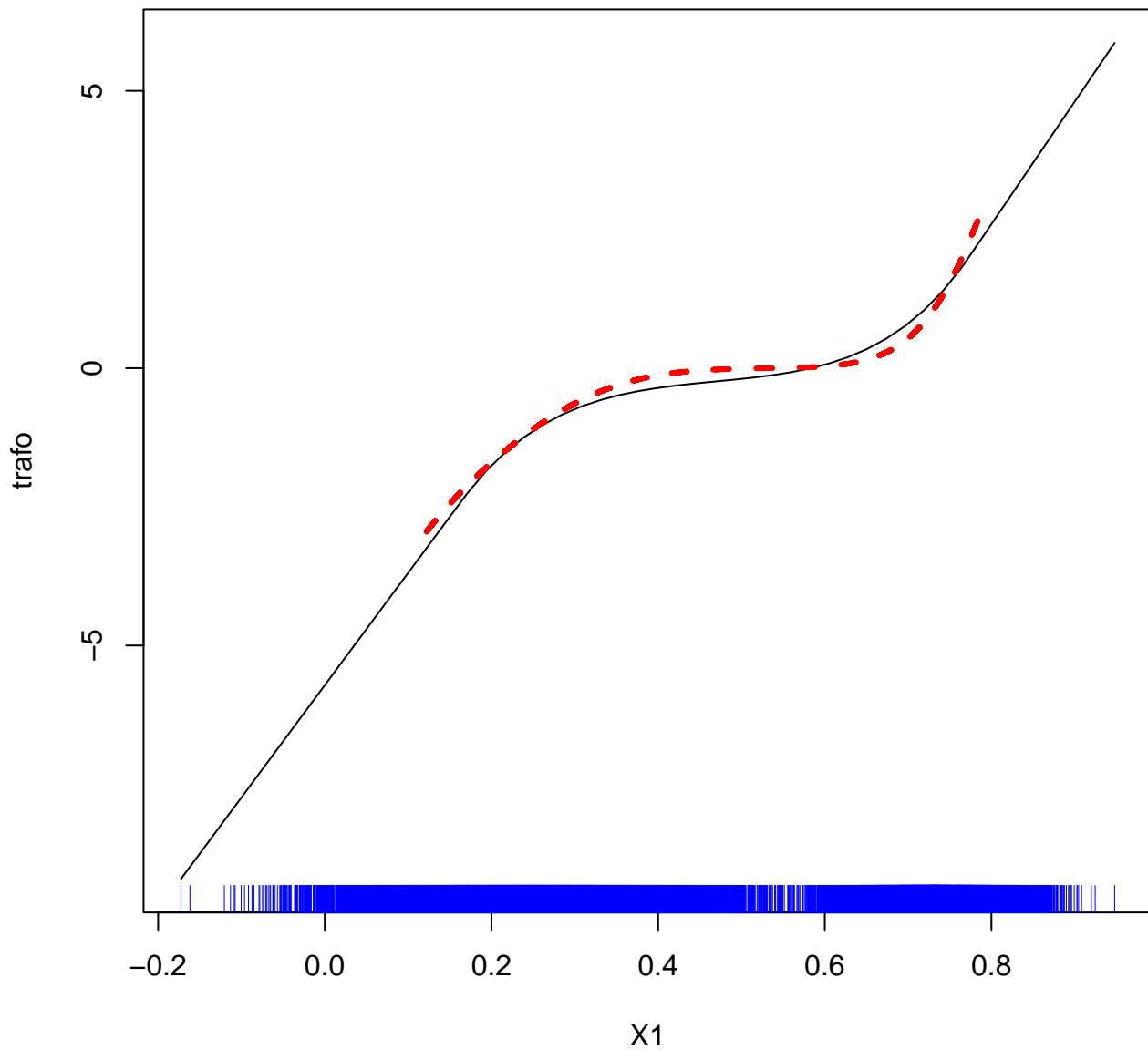


Last 50 epochs

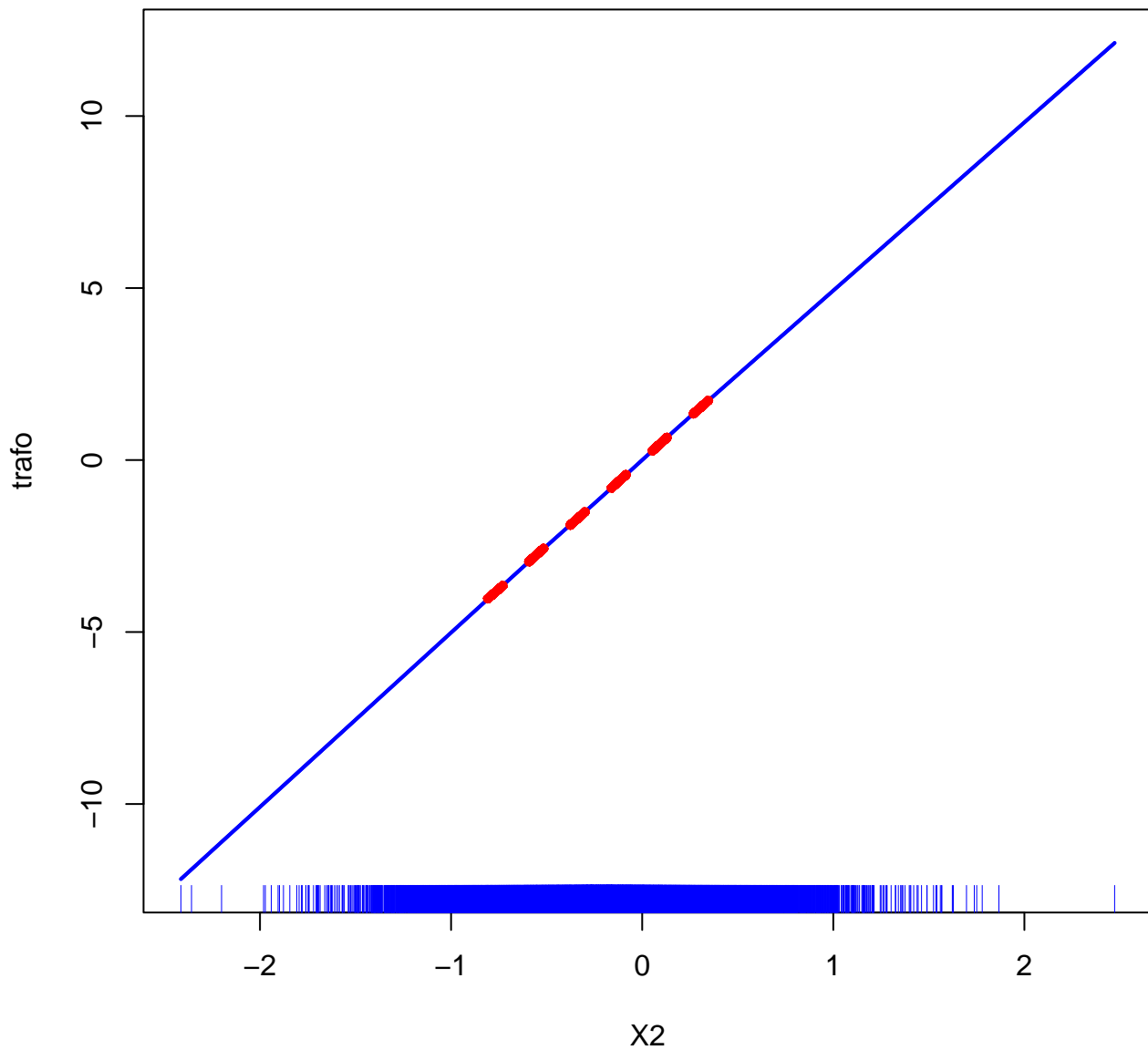




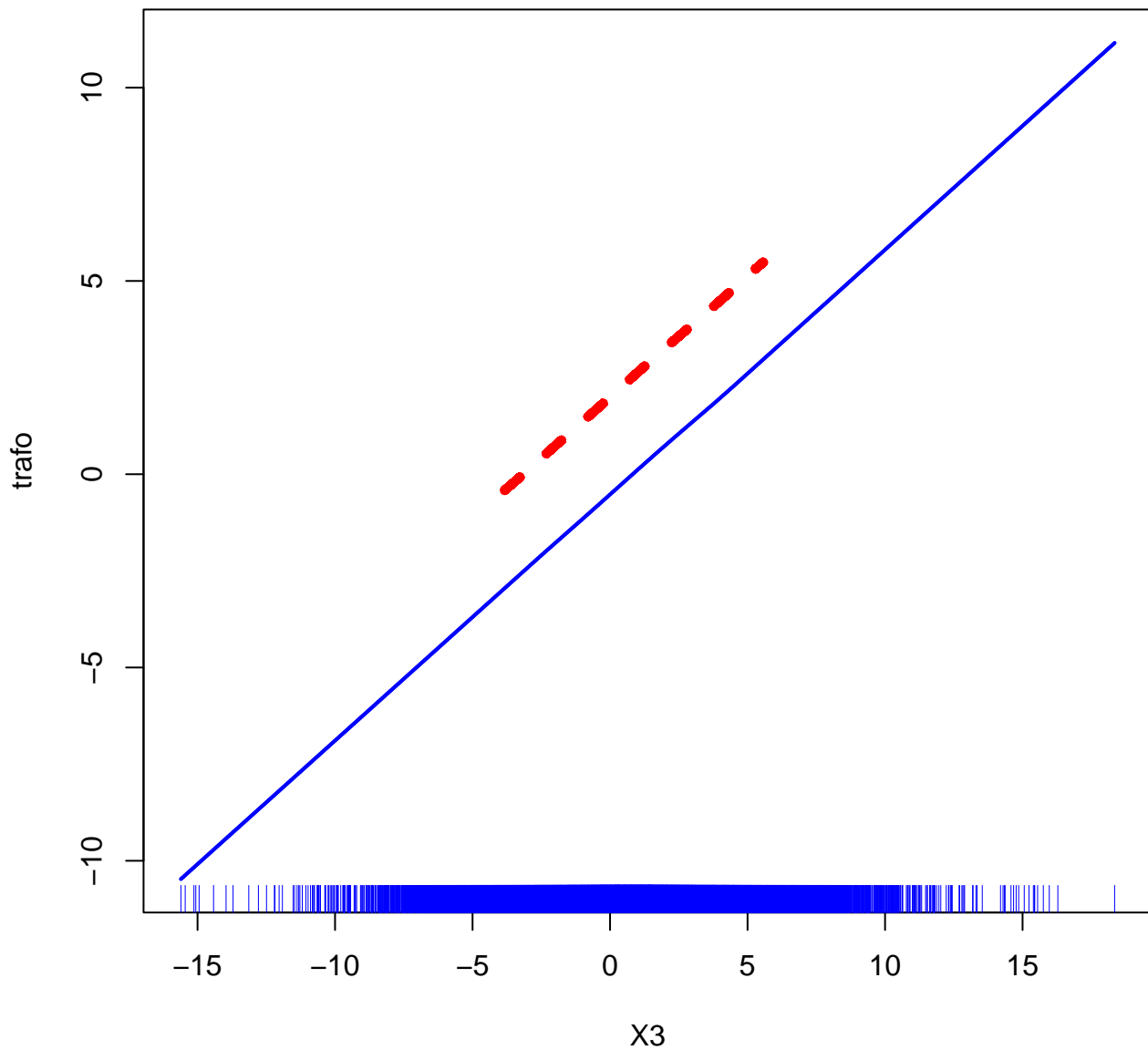
Black: COLR, Red: Our Model



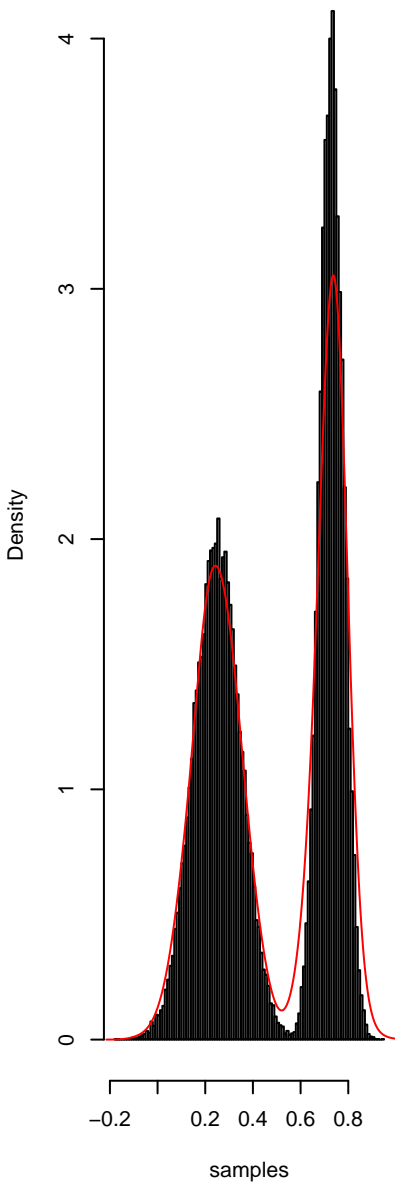
$h_I(X_2)$ Black: COLR, Red: Our Model



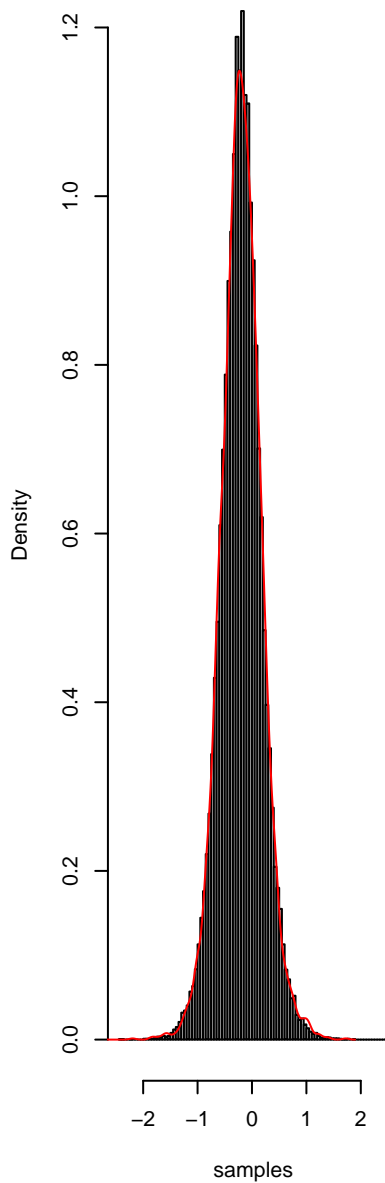
$h_I(X_3)$ Colr and Our Model



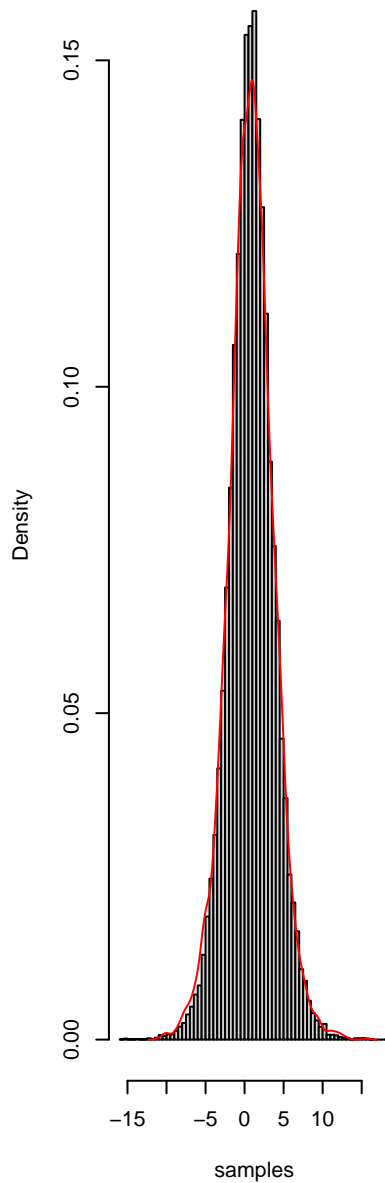
X1 red: ours, black: data



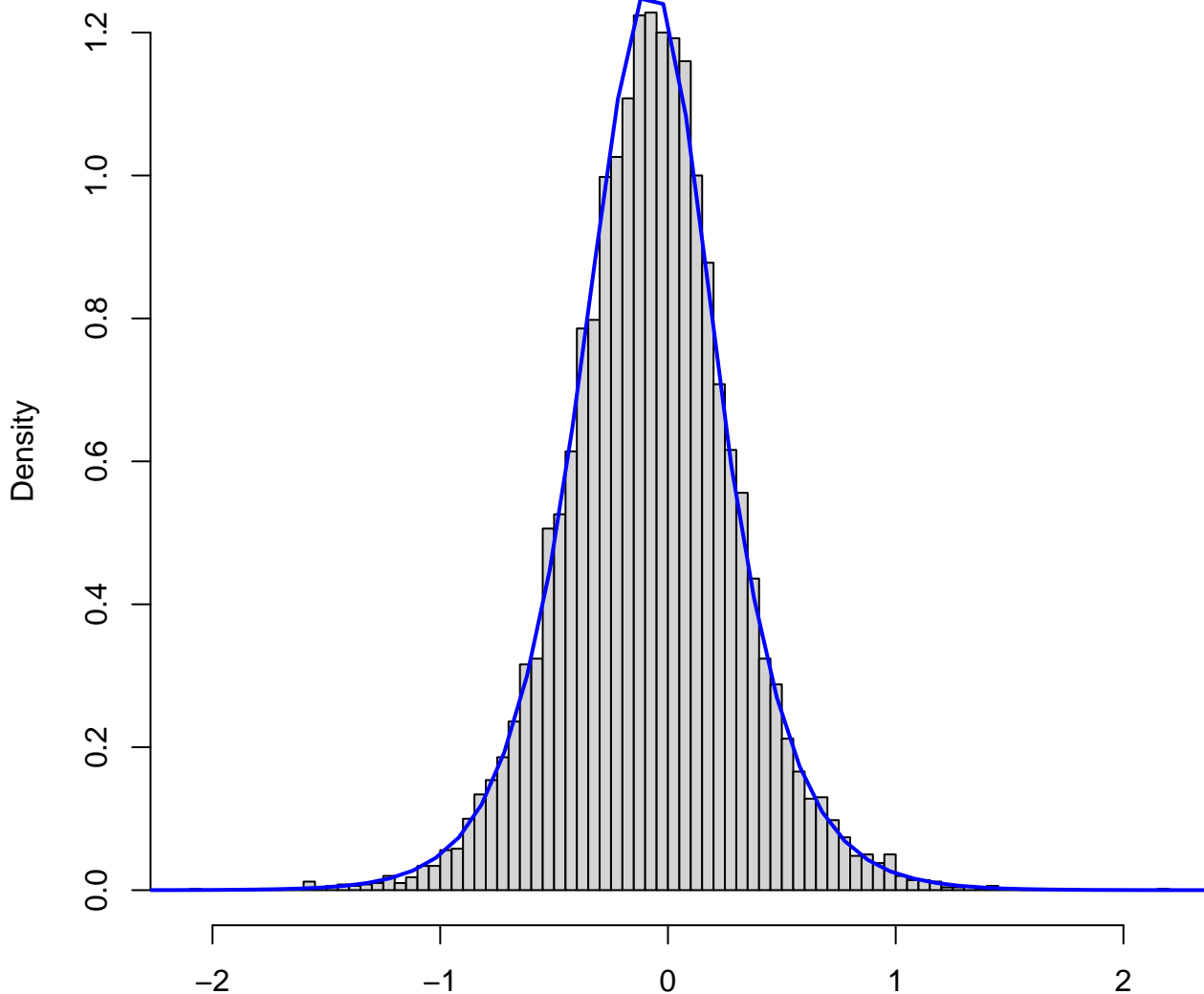
X2 red: ours, black: data



X3 red: ours, black: data

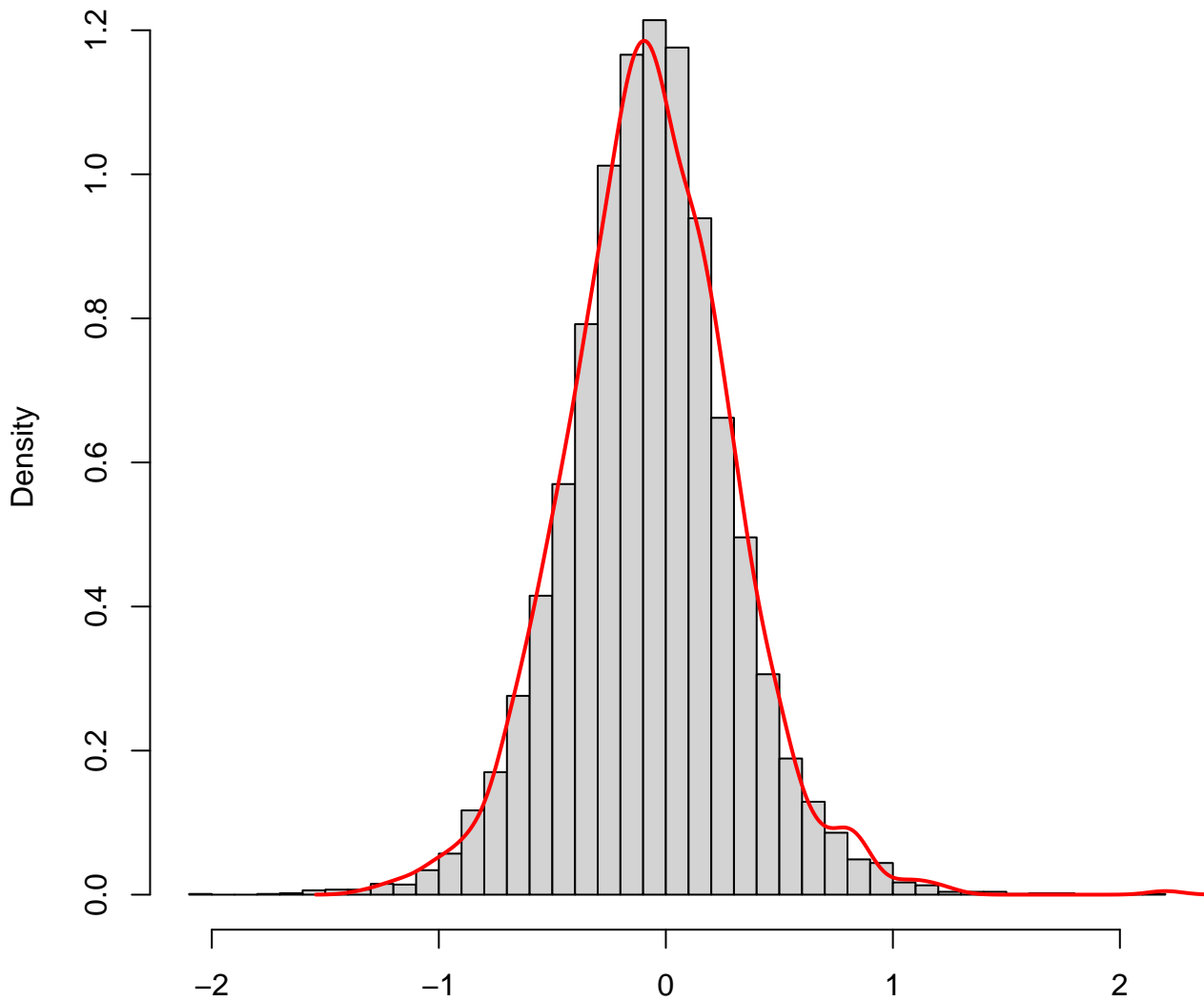


Do($X_1=0.2$) X_2



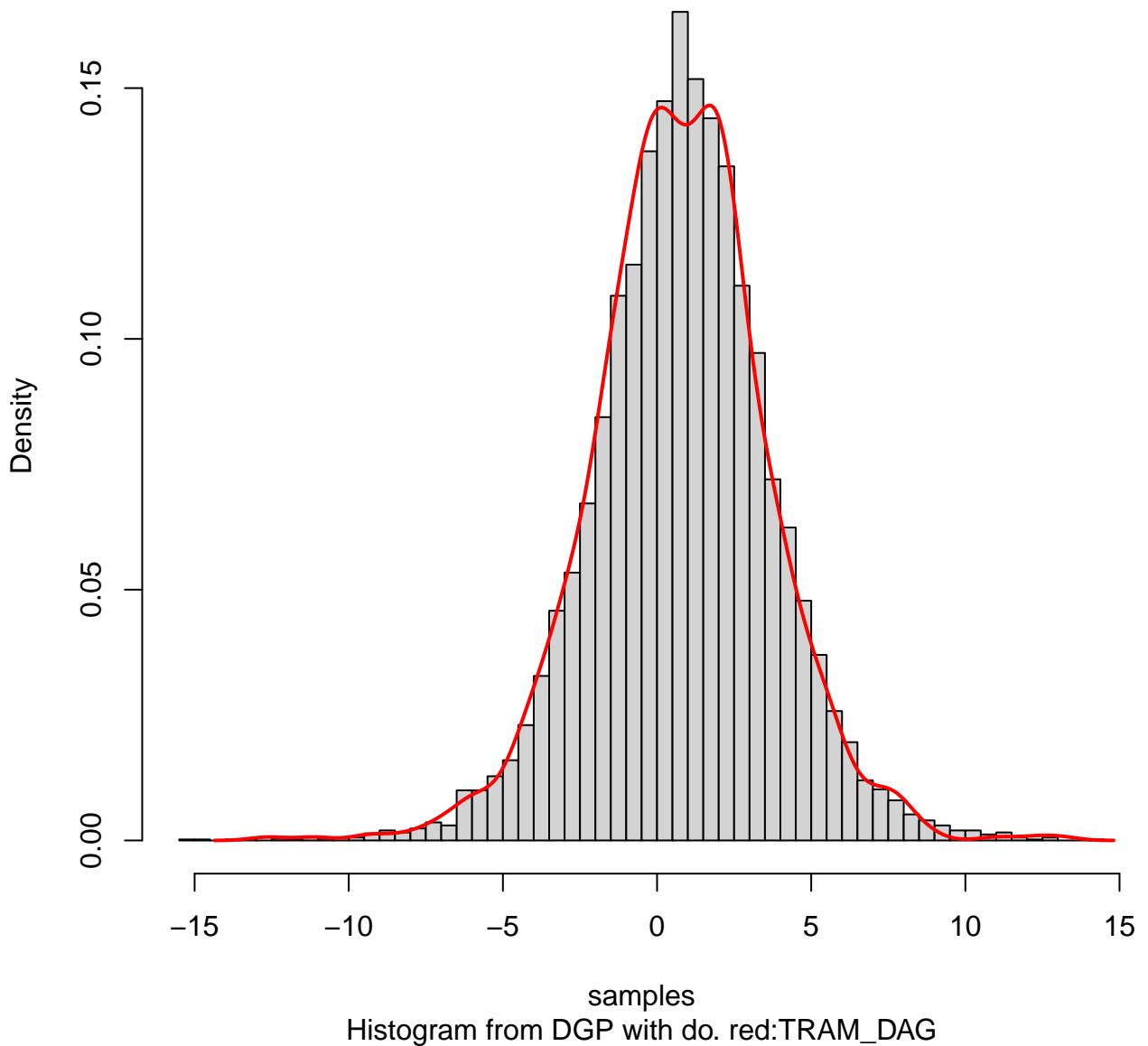
samples
Histogram from DGP with do. Blue: Colr

X2 | Do(X1=0.2)

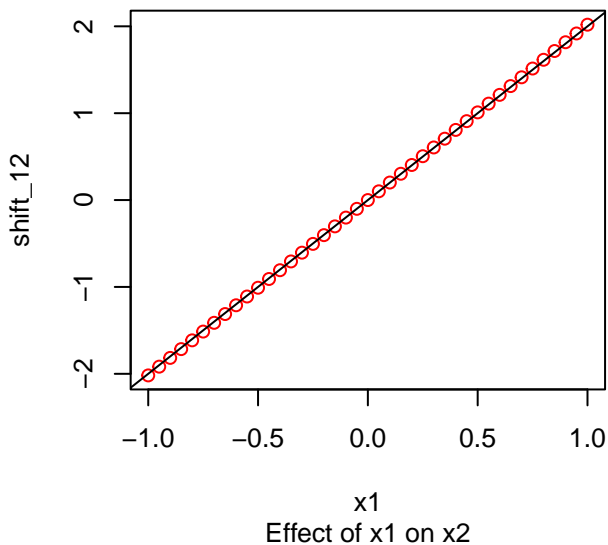


samples
Histogram from DGP with do. red:TRAM_DAG

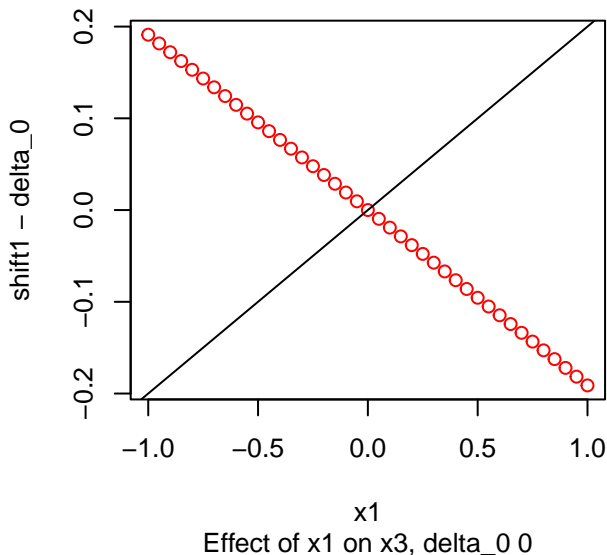
X3 | Do(X1=0.2)



LS-Term (black DGP, red Ours)



LS-Term (black DGP, red Ours)



CS-Term (black DGP $f_2(x)$, red Ours)

