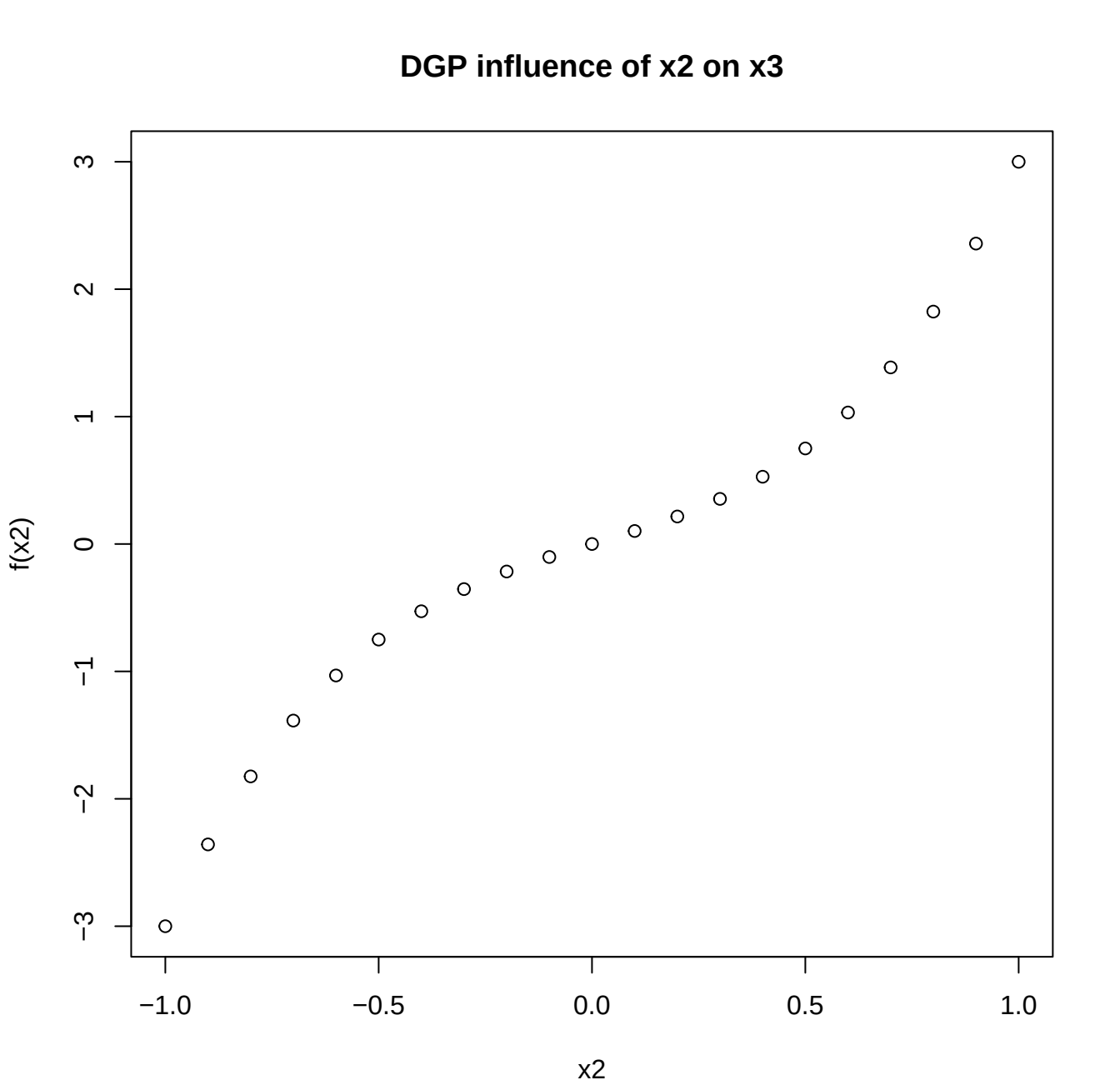
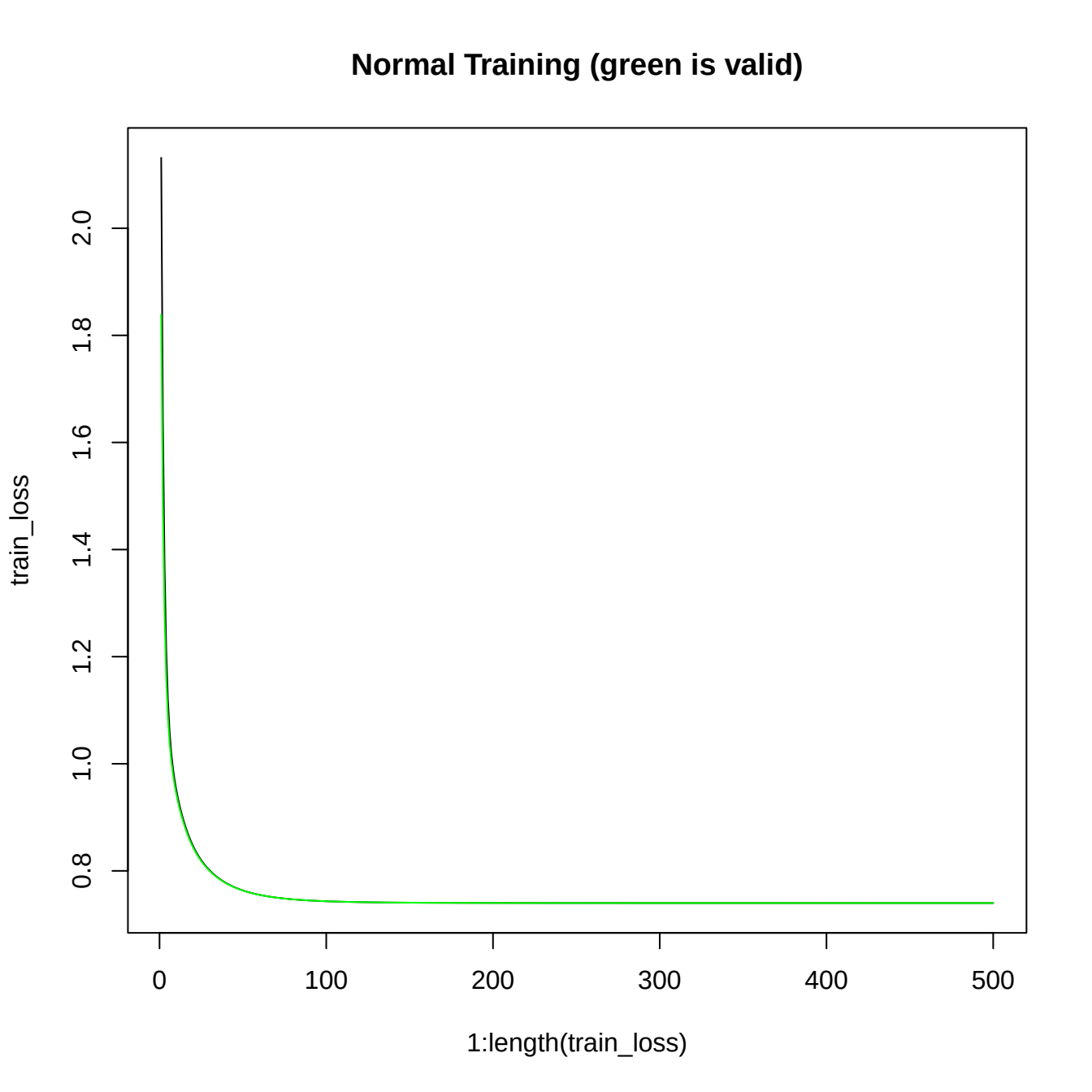
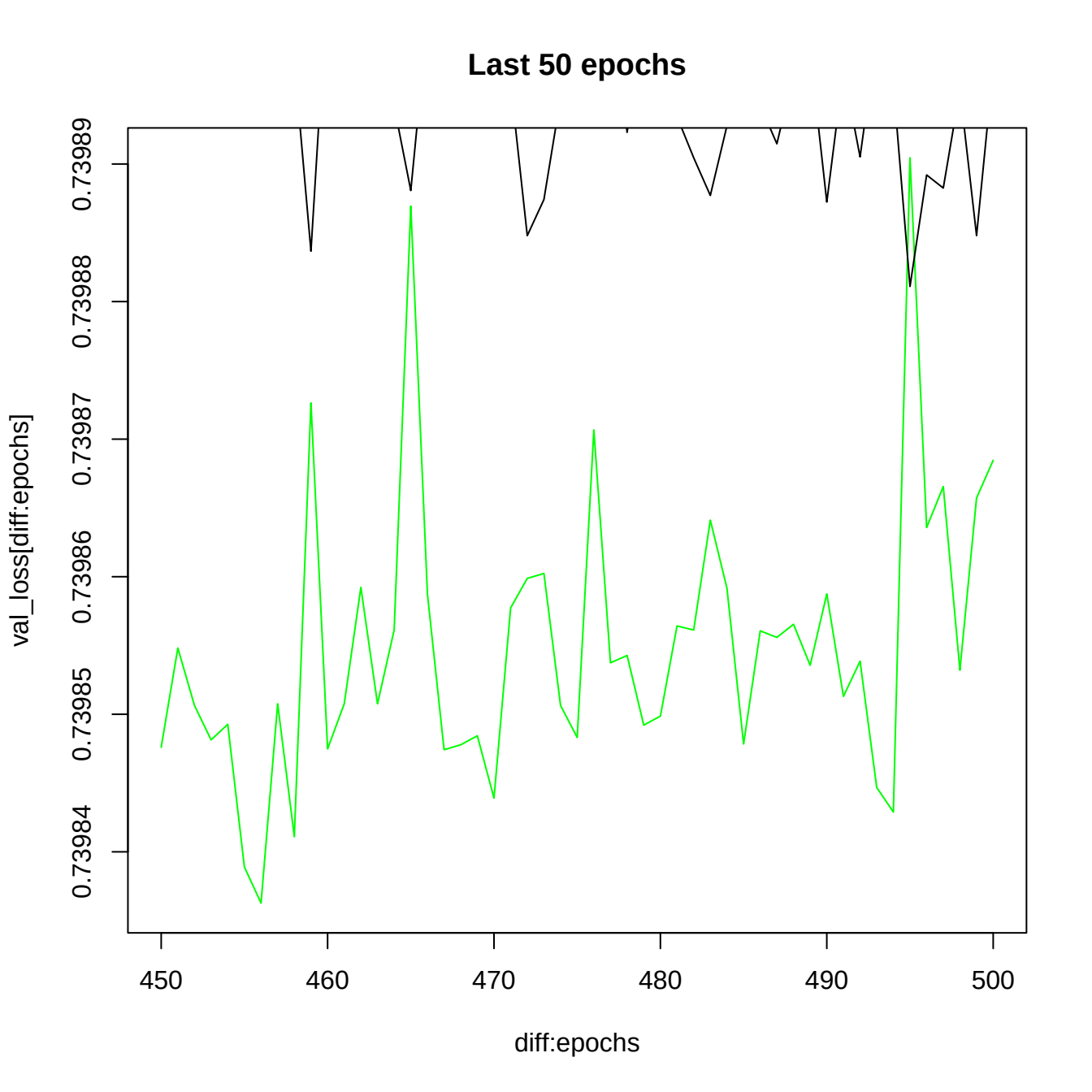
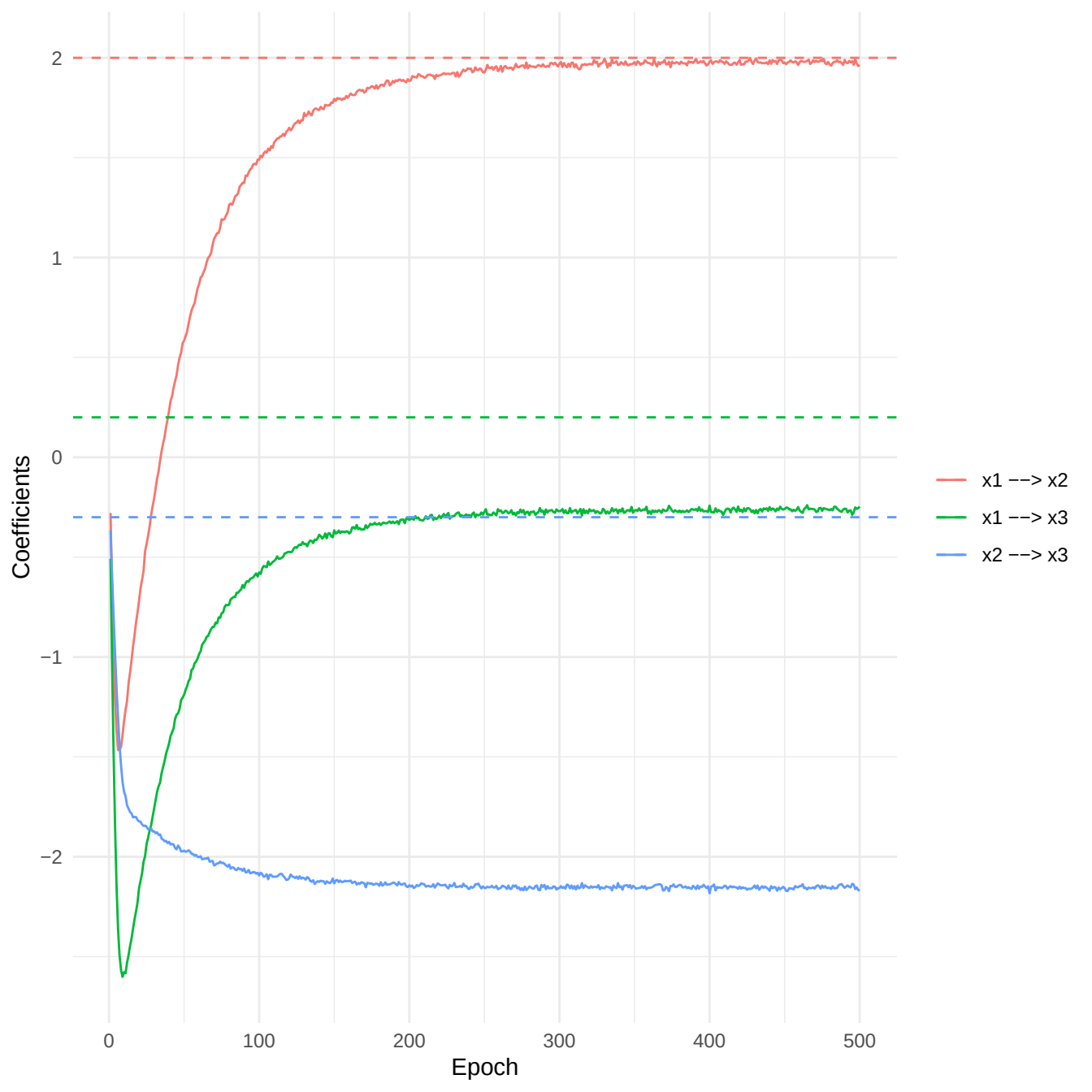
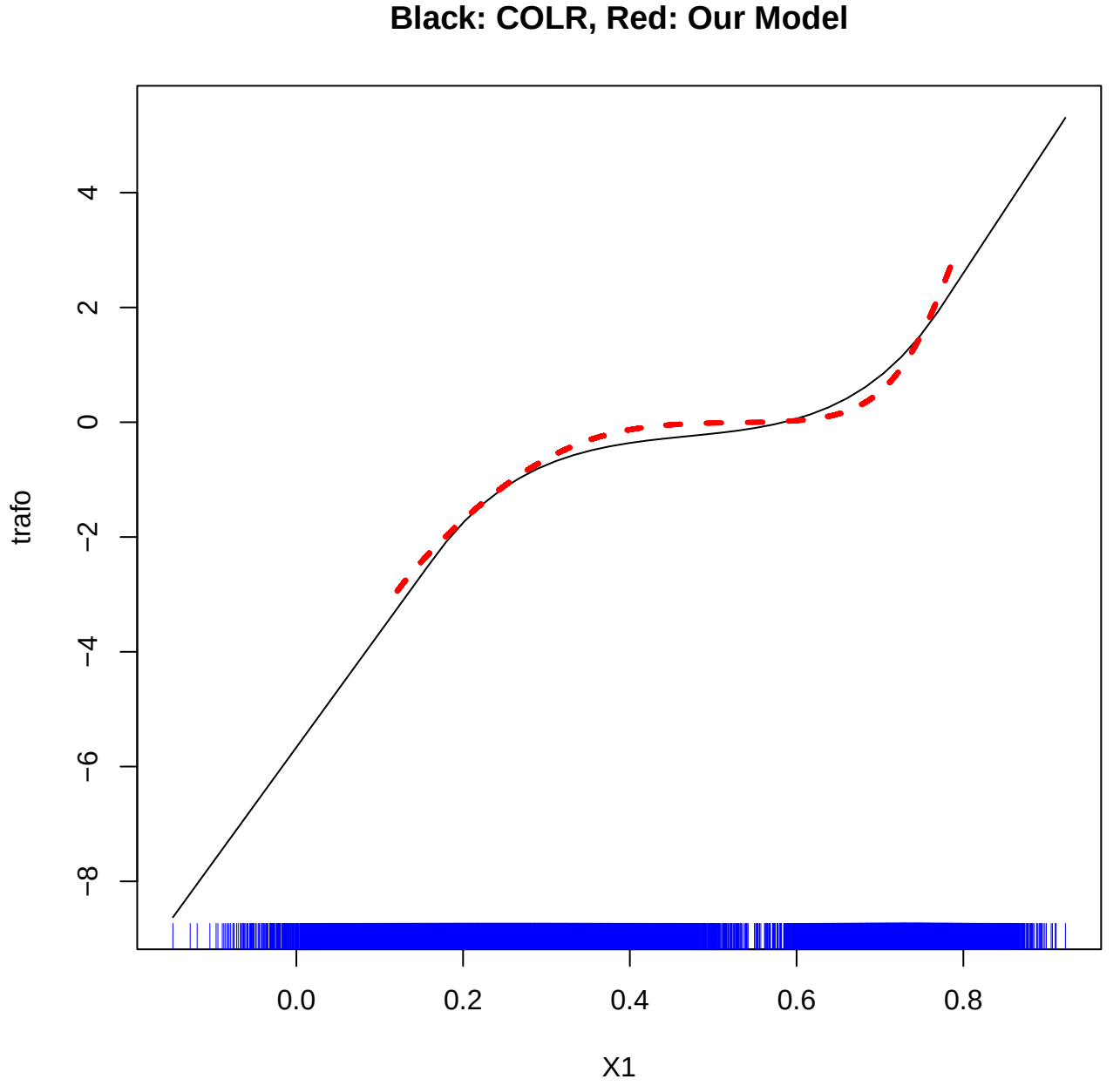
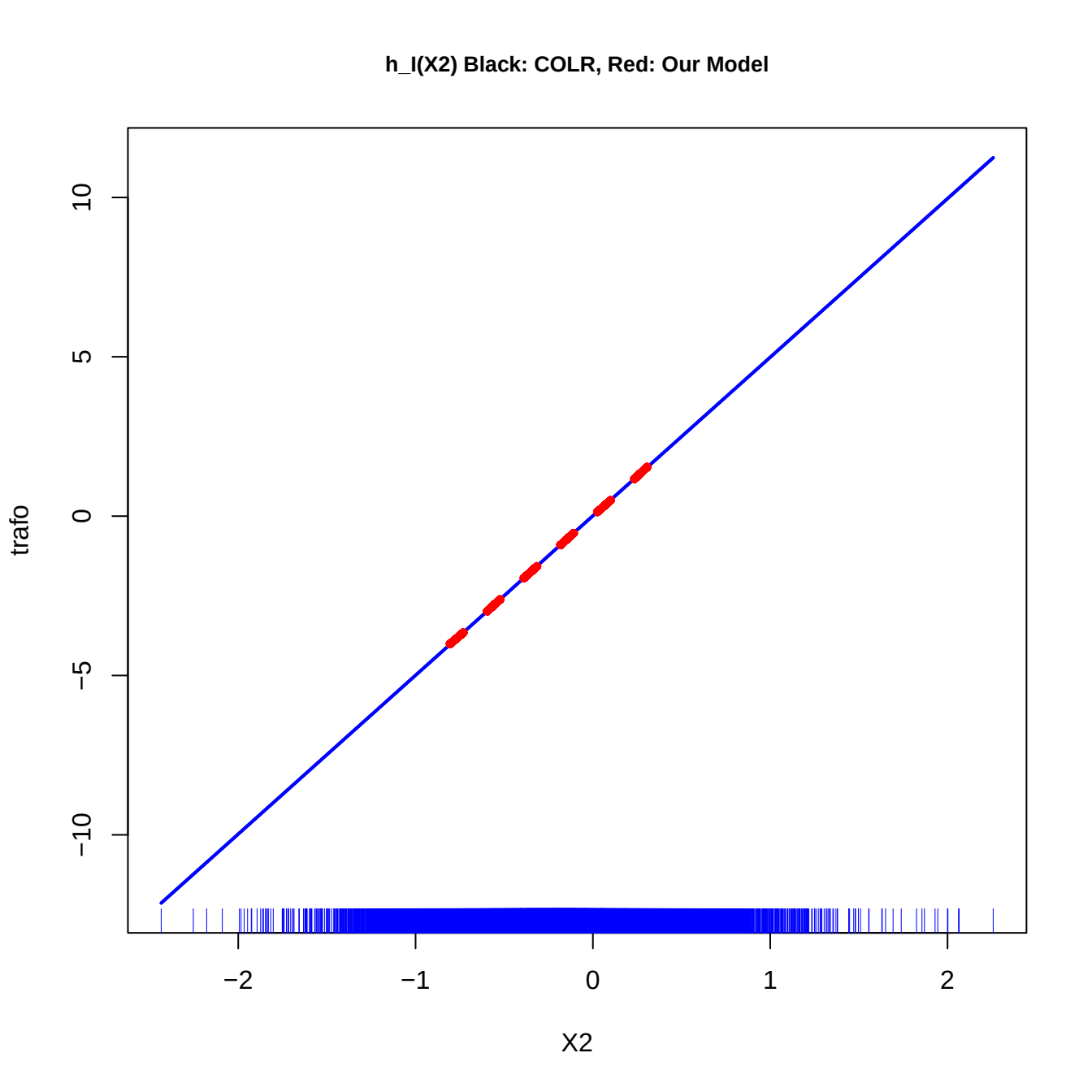

DGP influence of x2 on x3
f(x2)
3
2
1
0
−1
−2
−3
−1.0
−0.5
0.0
0.5
1.0
x2
summerof24/runs/triangle_structured_continous/run_nodes25/triangle_mixed_DPG2x3+x_Mod

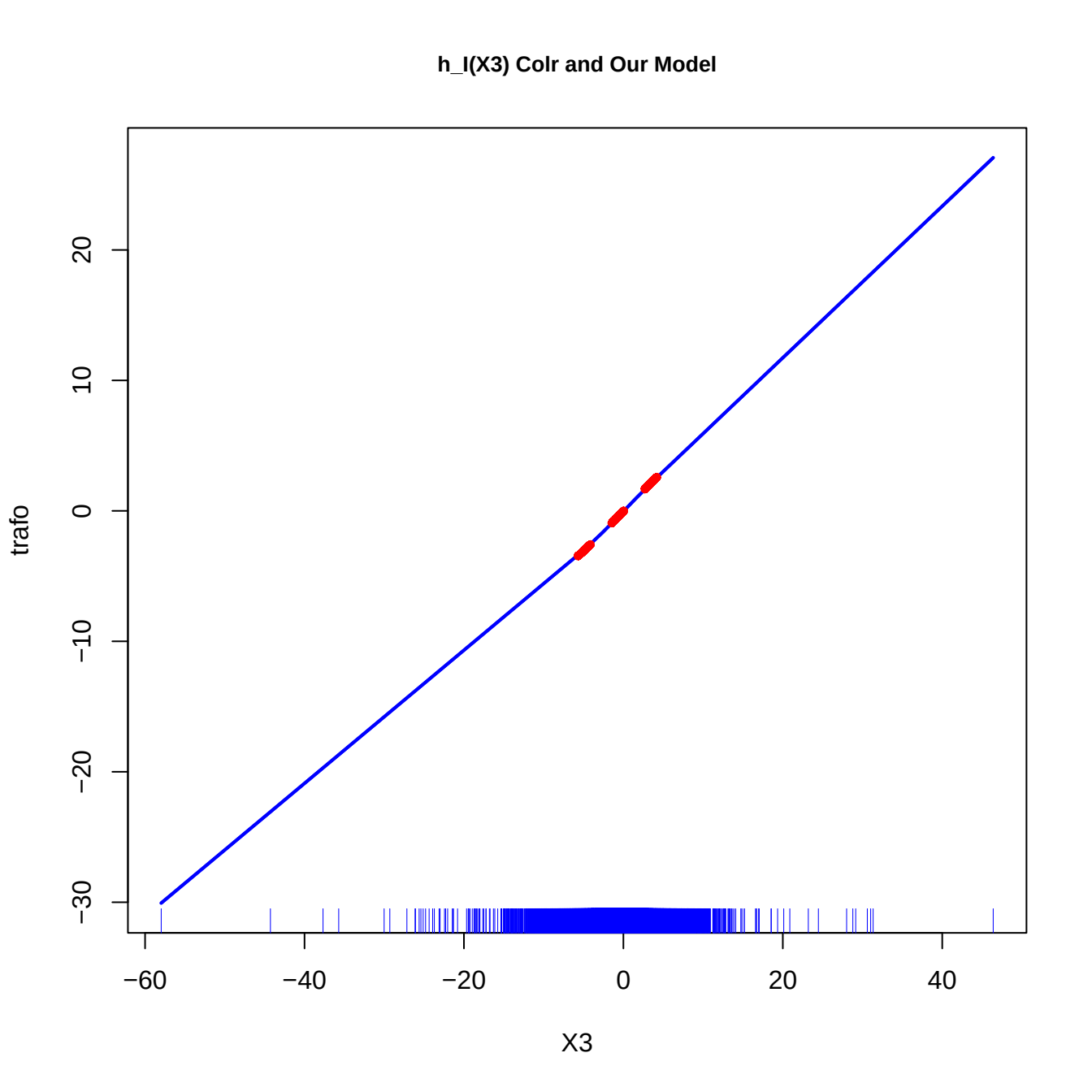

Normal Training (green is valid)
train_loss
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0
100
200
300
400
500
1:length(train_loss)

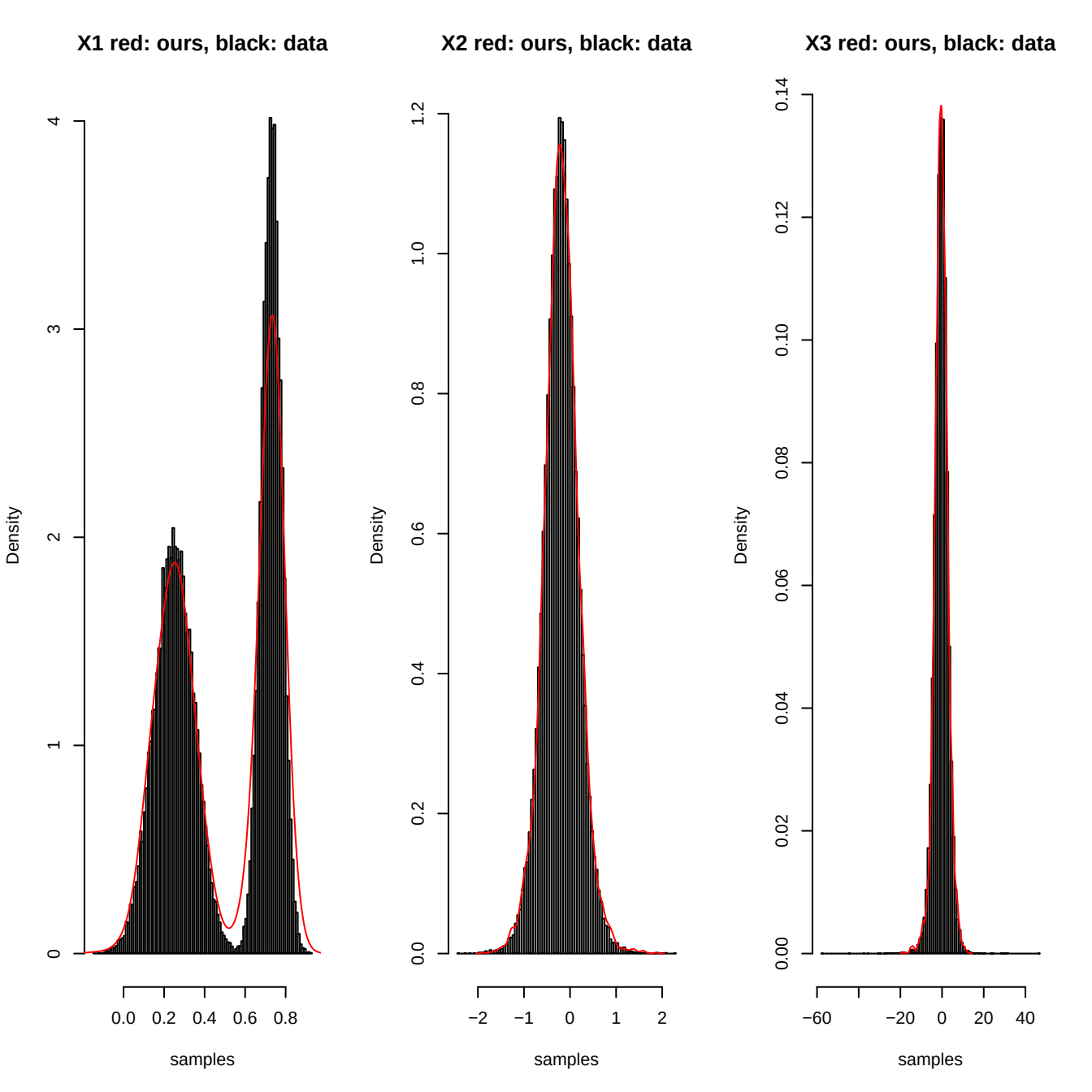

Last 50 epochs
val_loss[diff:epochs]
0.73984
0.73985
0.73986
0.73987
0.73988
0.73989
450
460
470
480
490
500
diff:epochs

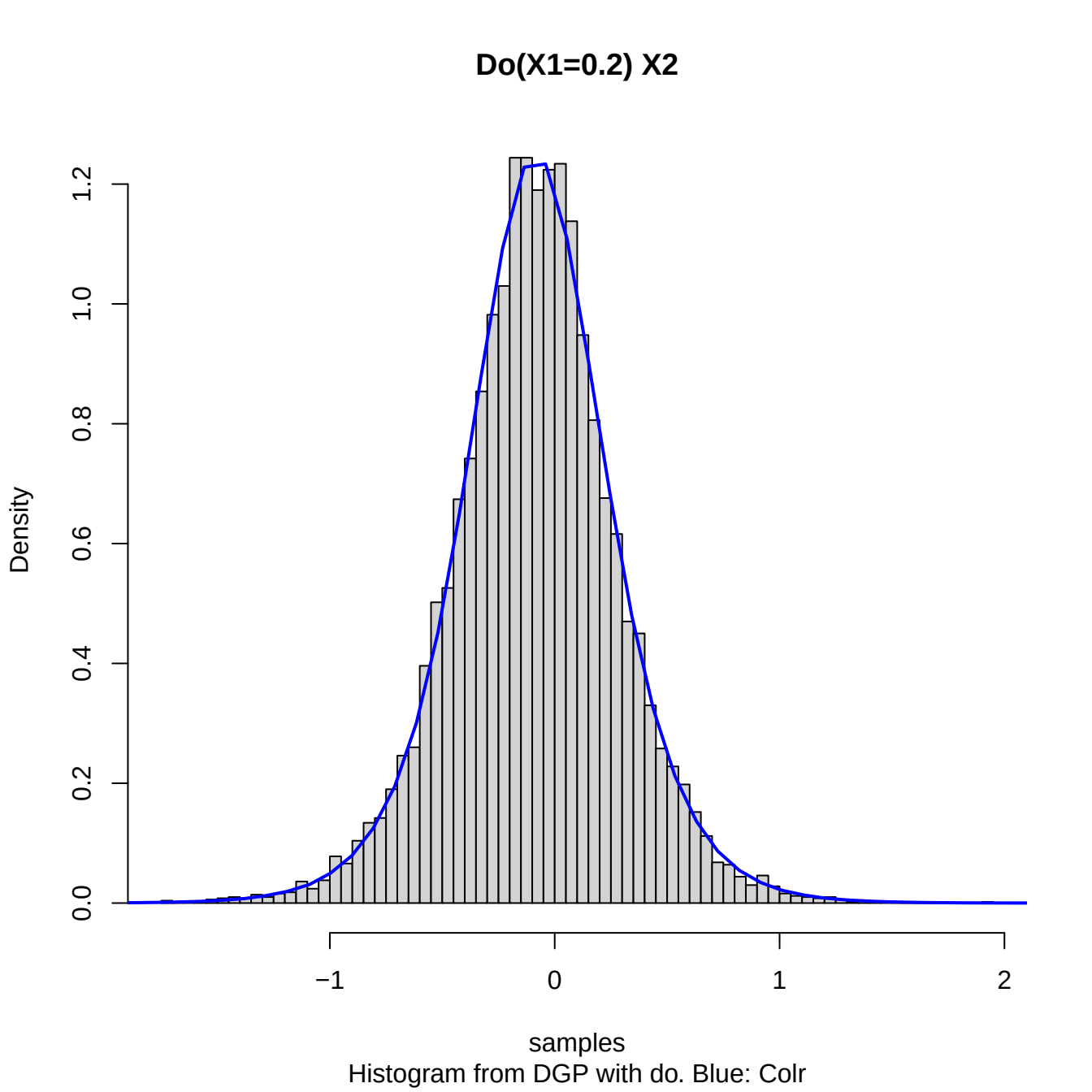

2
1
0
−1
−2
Coefficients
0
100
200
300
400
500
Epoch
x1 −−> x2
x1 −−> x3
x2 −−> x3

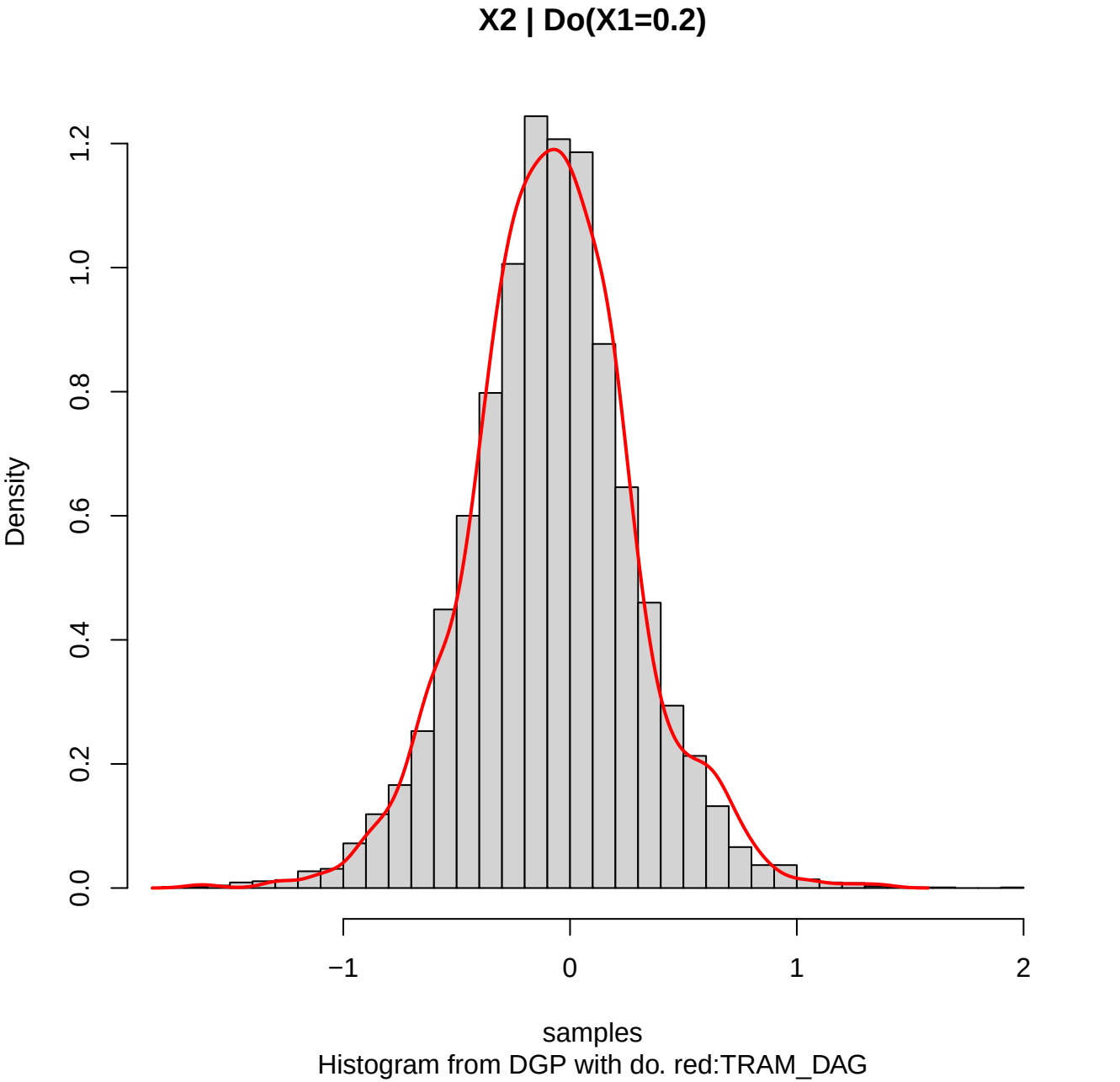

Black: COLR, Red: Our Model
trafo
4
2
0
−2
−4
−6
−8
0.0
0.2
0.4
0.6
0.8
X1

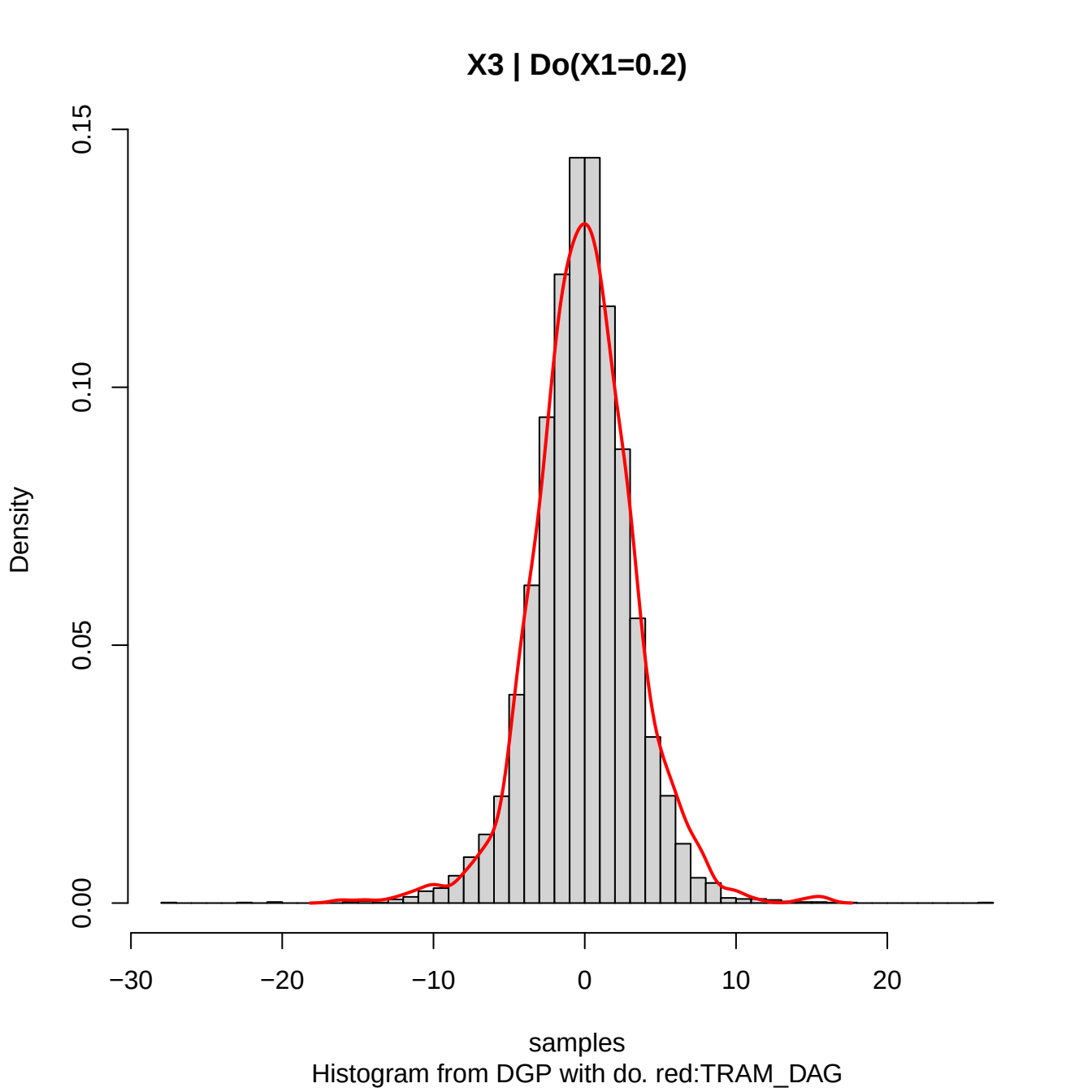

h_l(X2) Black: COLR, Red: Our Model
trafo
10
5
0
−5
−10
−2
−1
0
1
2
X2

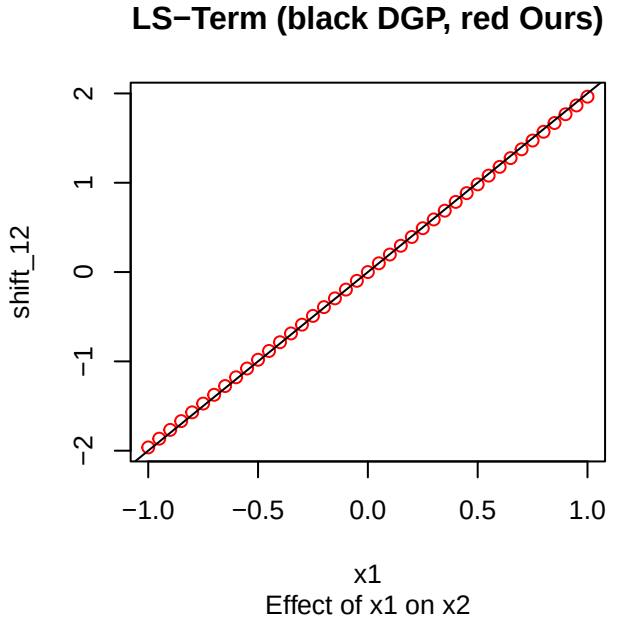

h_I(X3) Colr and Our Model
trafo
20
10
0
−10
−20
−30
−60
−40
−20
0
20
40
X3

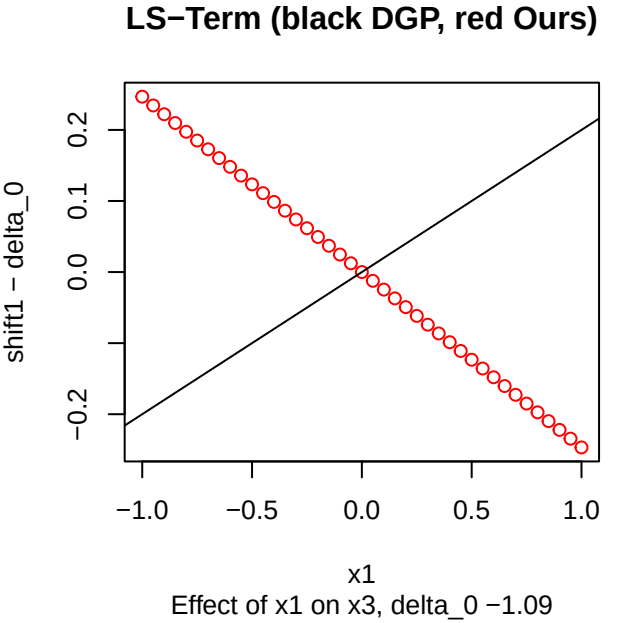

X1 red: ours, black: data
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
samples
X2 red: ours, black: data
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−2
−1
0
1
2
samples
X3 red: ours, black: data
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
−60
−20
0
20
40
samples

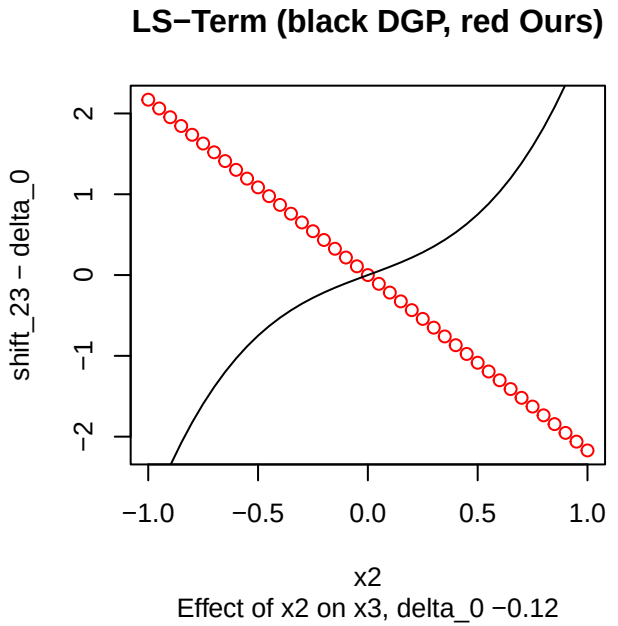

Do(X1=0.2) X2
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−1
0
1
2
samples
Histogram from DGP with do. Blue: Colr


X2 | Do(X1=0.2)
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−1
0
1
2
samples
Histogram from DGP with do. red:TRAM_DAG


X3 | Do(X1=0.2)
Density
0.00
0.05
0.10
0.15
−30
−20
−10
0
10
20
samples
Histogram from DGP with do. red:TRAM_DAG


LS−Term (black DGP, red Ours)
shift_12
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
x1
Effect of x1 on x2


LS−Term (black DGP, red Ours)
shift1 − delta_0
0.2
0.1
0.0
−0.2
−1.0
−0.5
0.0
0.5
1.0
x1
Effect of x1 on x3, delta_0 −1.09


LS−Term (black DGP, red Ours)
shift_23 − delta_0
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0
x2
Effect of x2 on x3, delta_0 −0.12