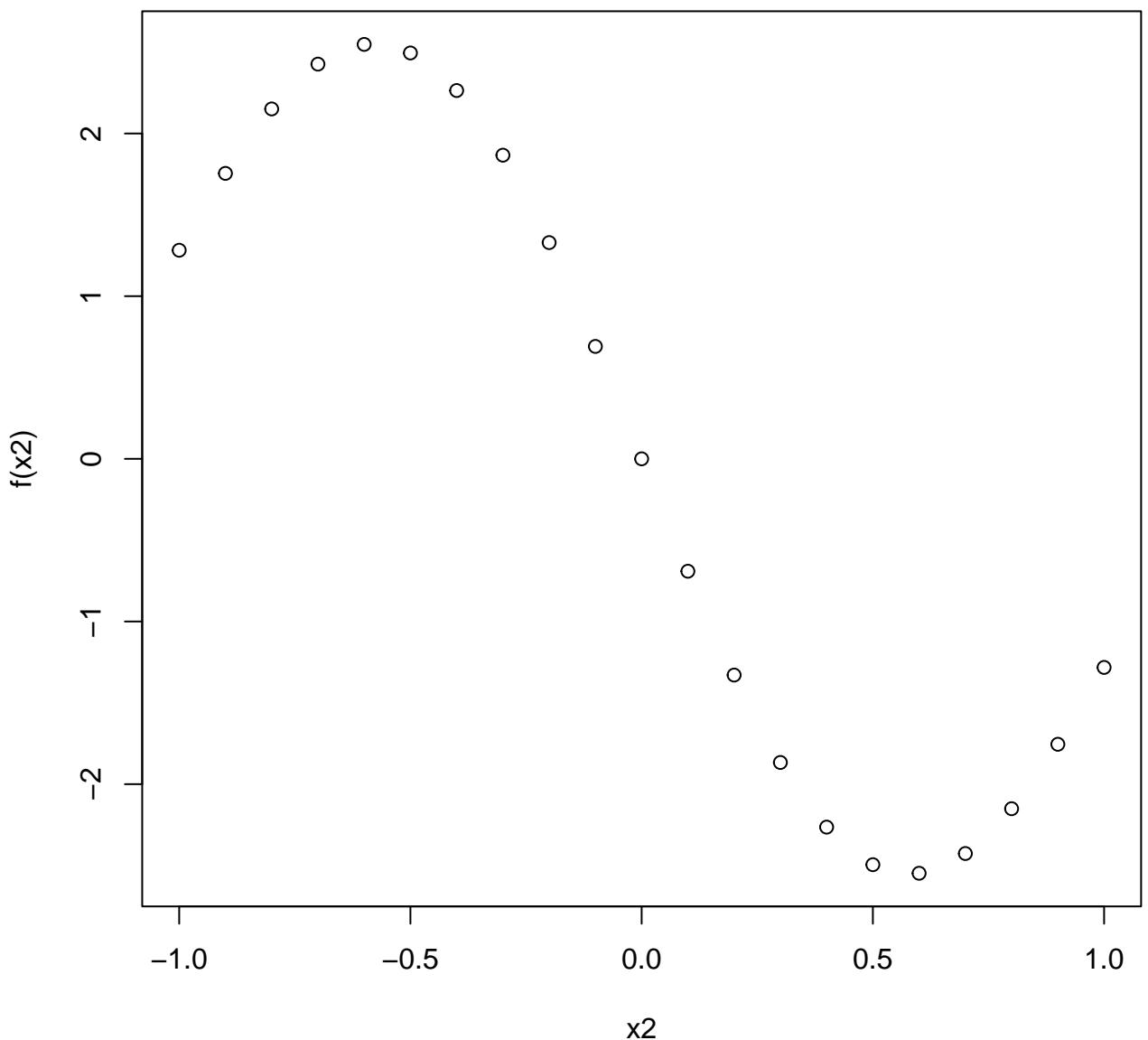
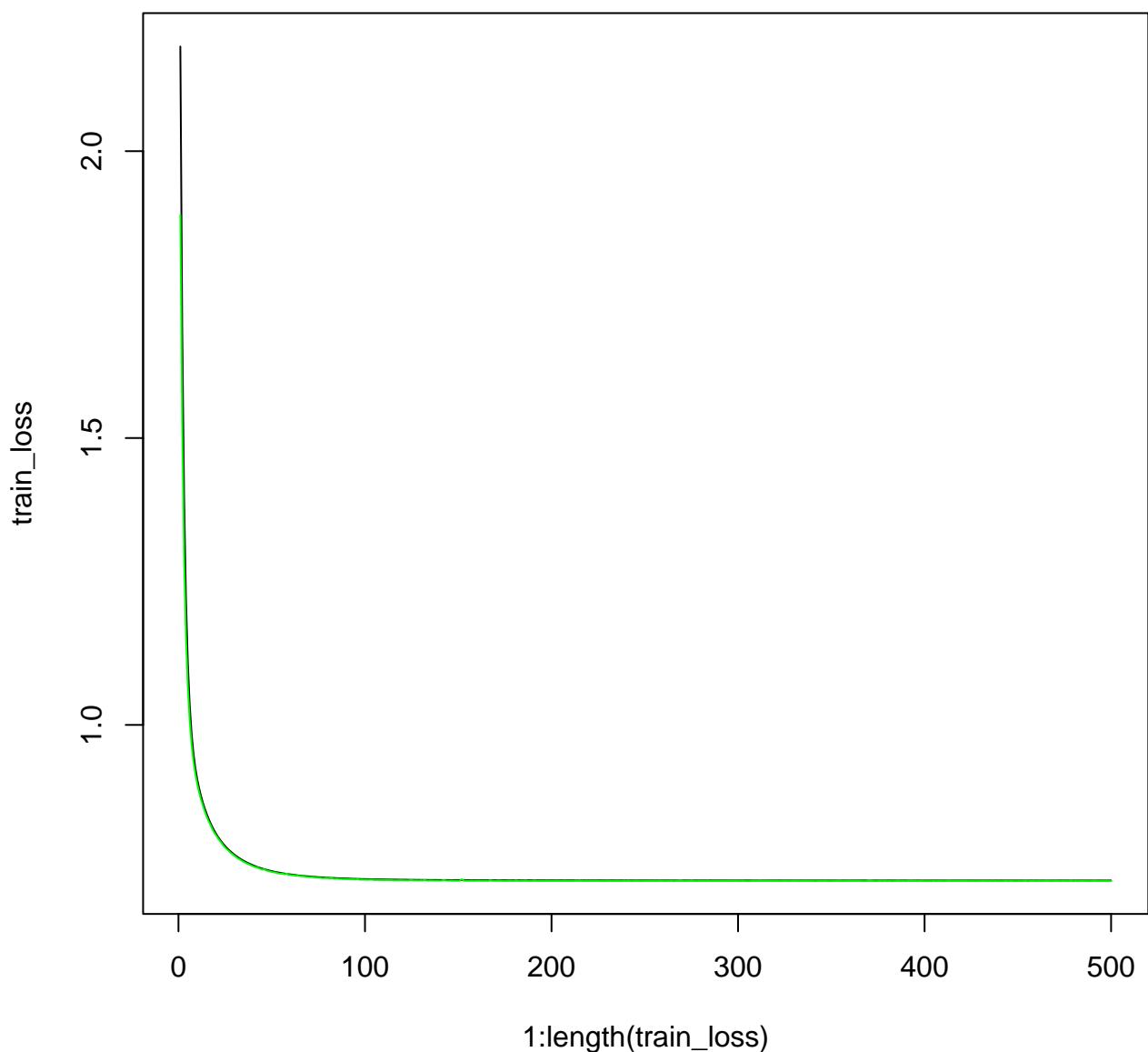


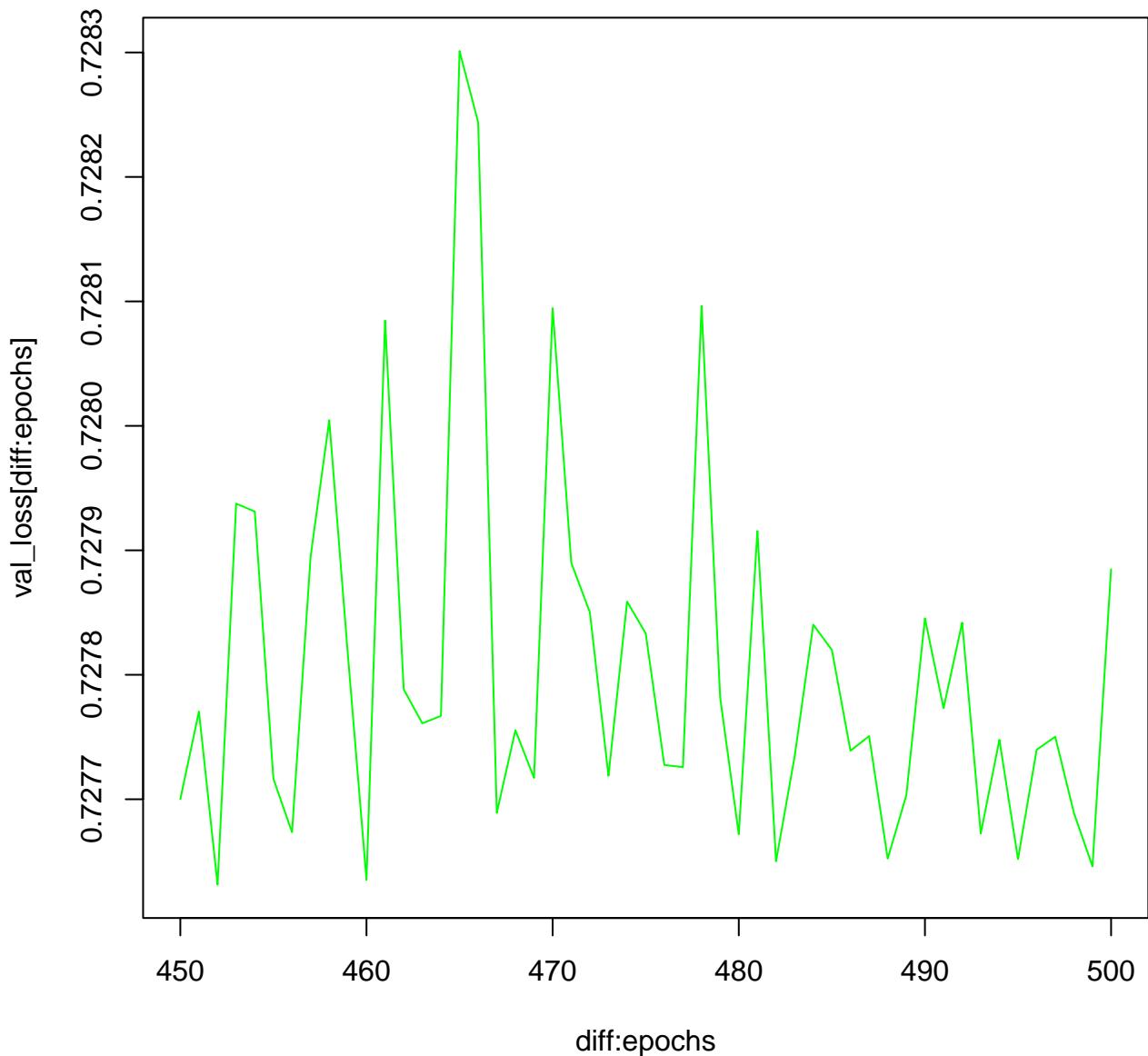
## DGP influence of $x_2$ on $x_3$

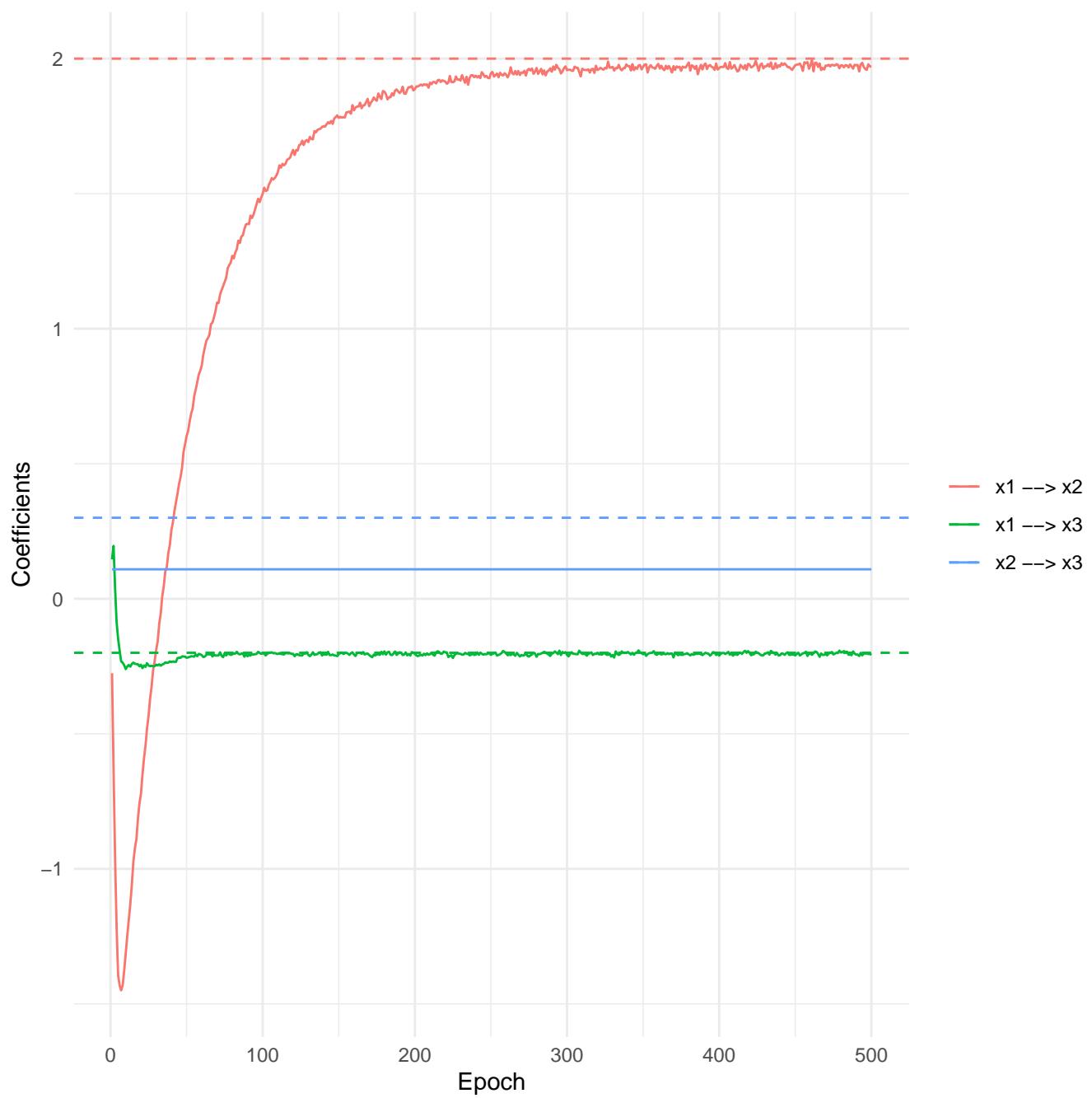


**Training (black: train, green: valid)**

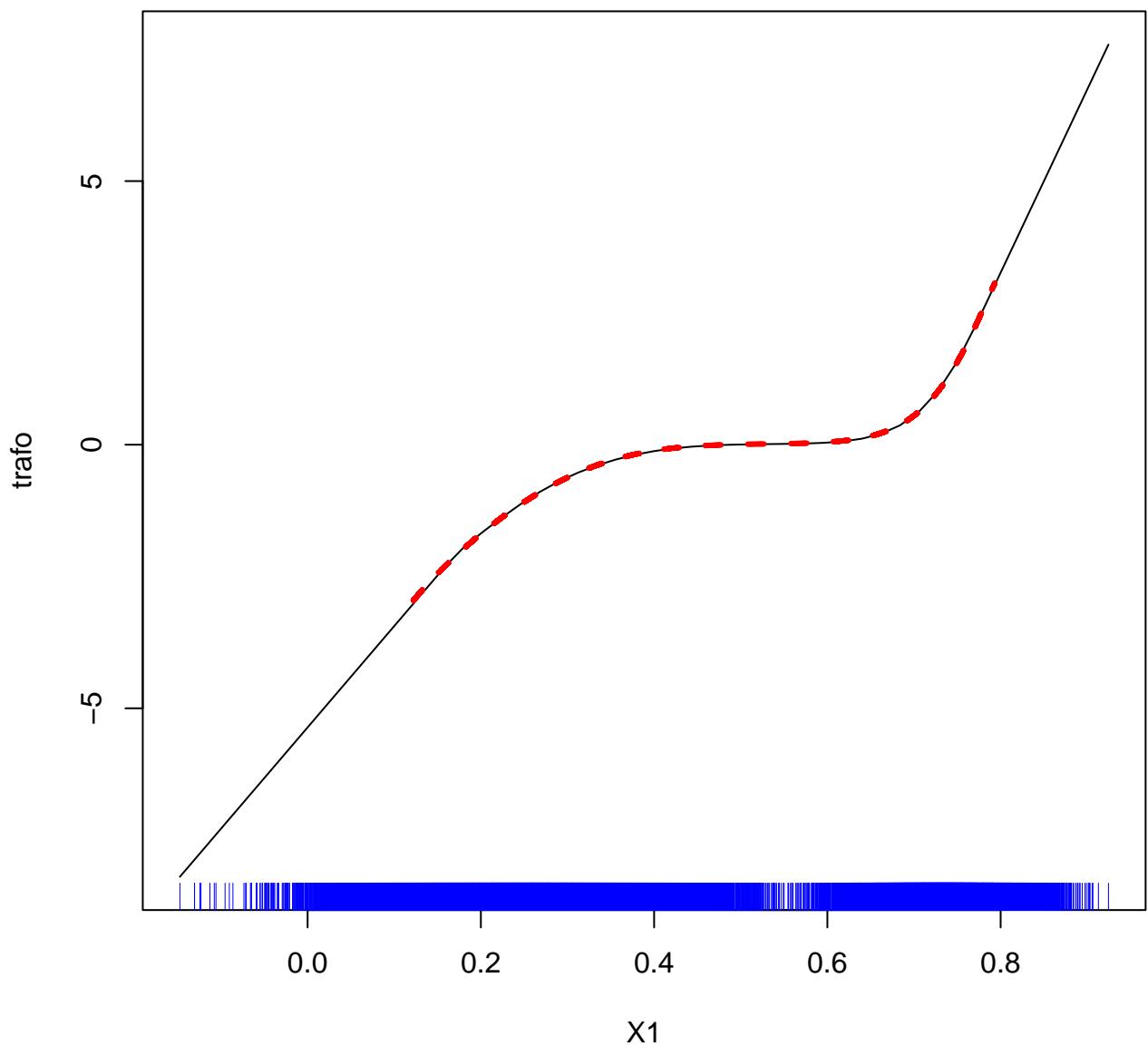


## Last 50 epochs

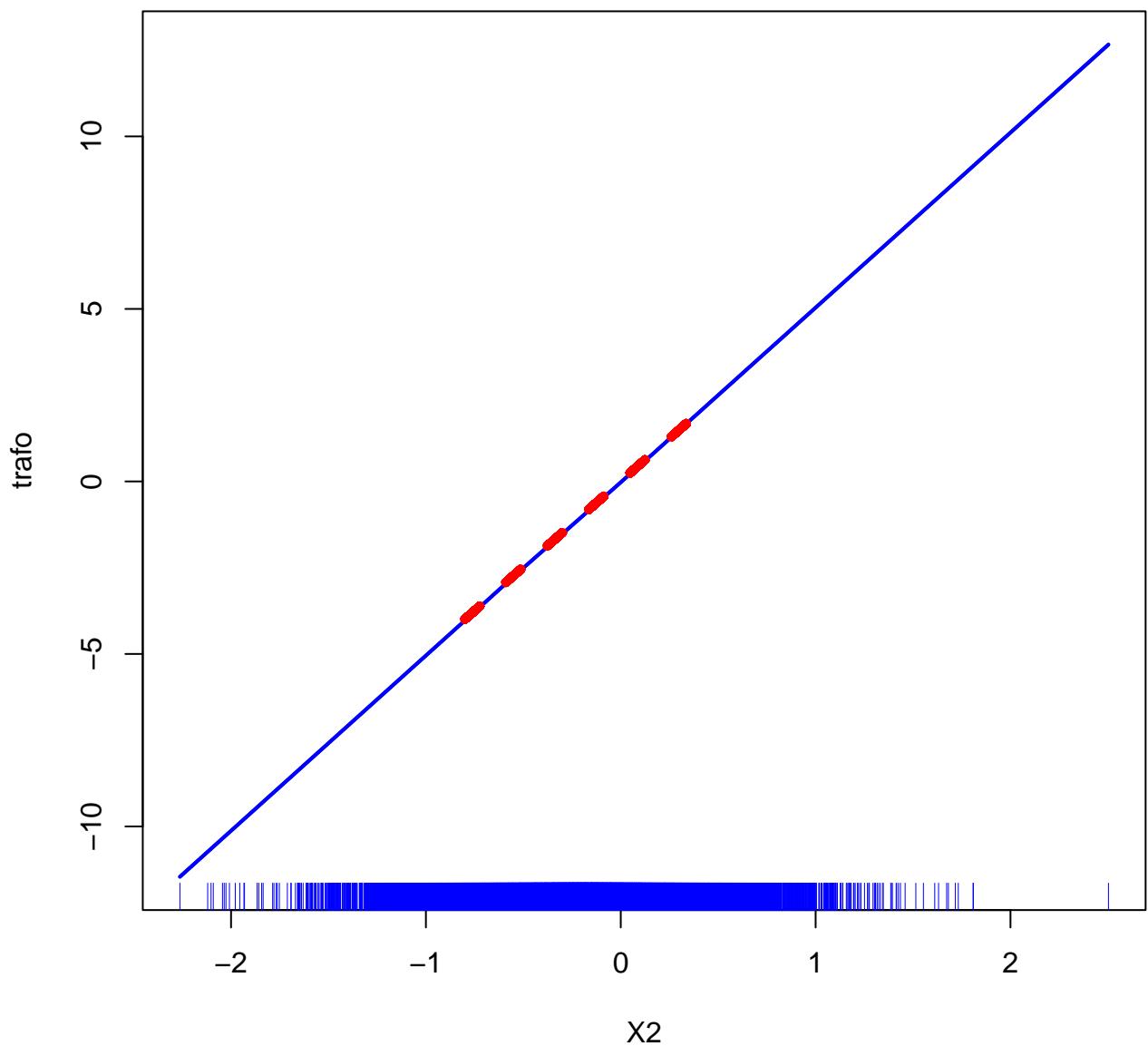




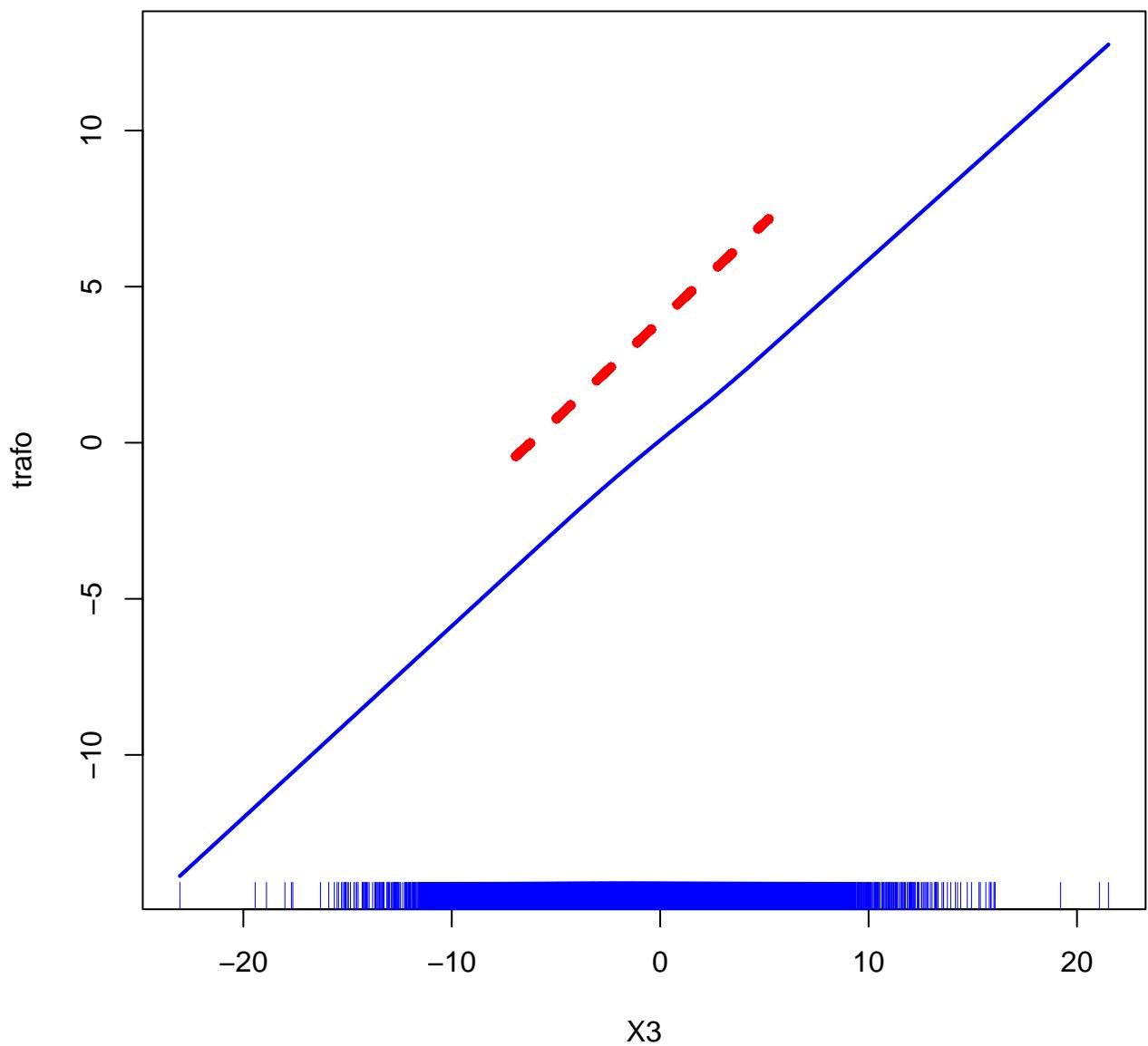
**Black: COLR, Red: Our Model**



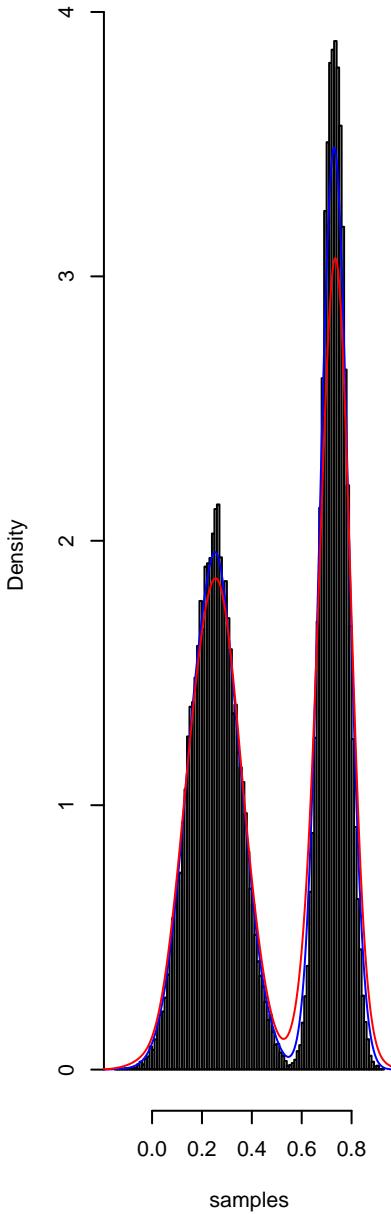
$h_I(X_2)$  Black: COLR, Red: Our Model



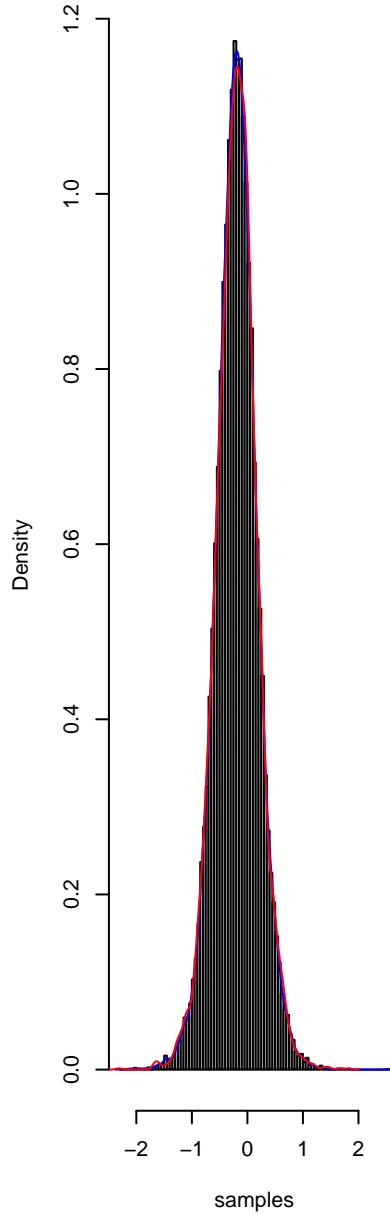
**h\_I(X3) Colr and Our Model**



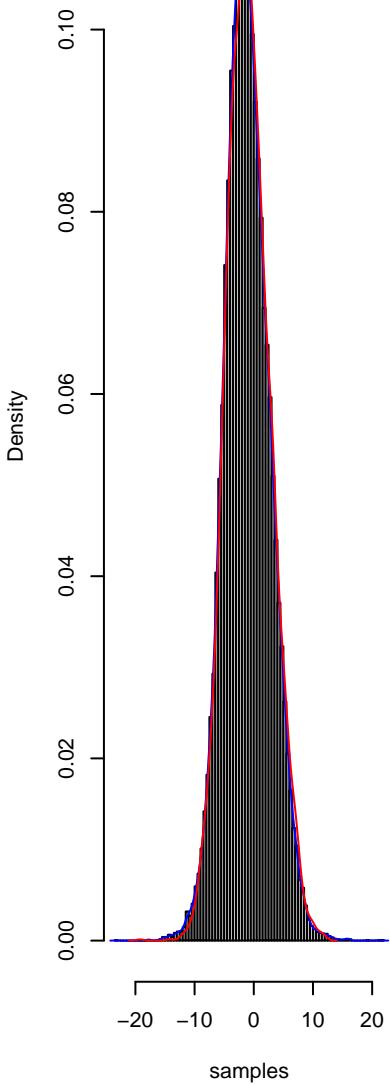
X1 red: ours, black: data



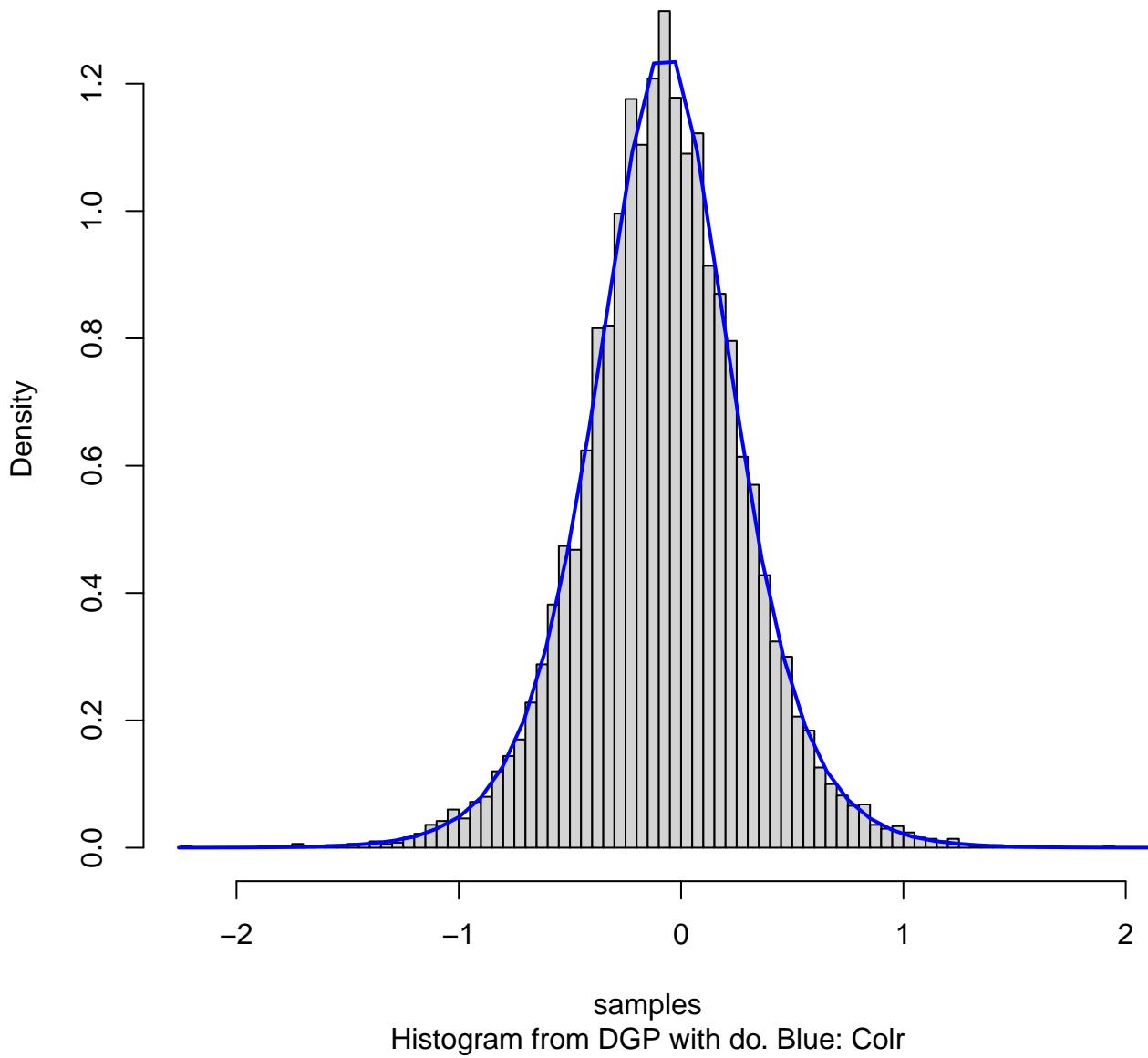
X2 red: ours, black: data



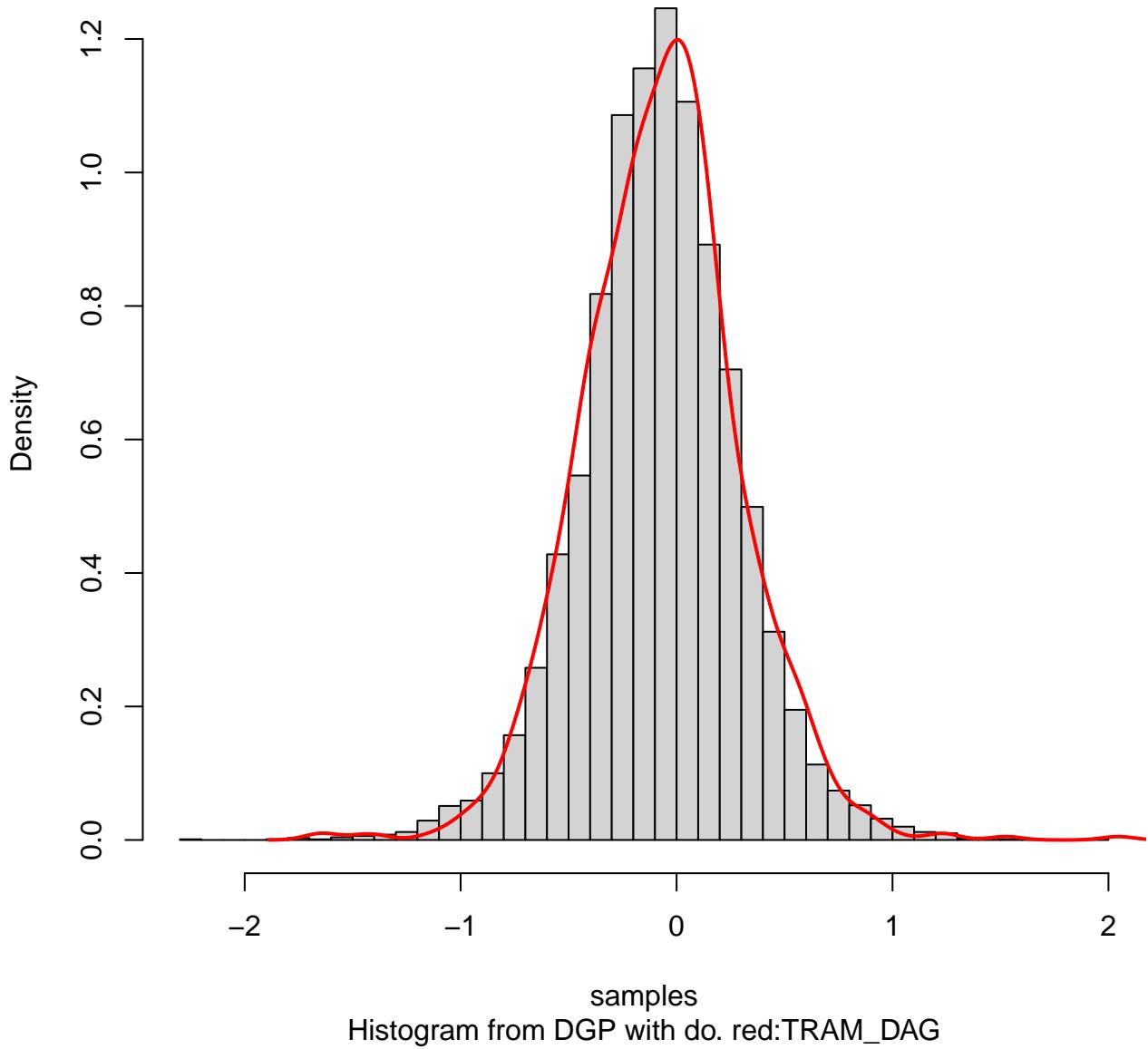
X3 red: ours, black: data



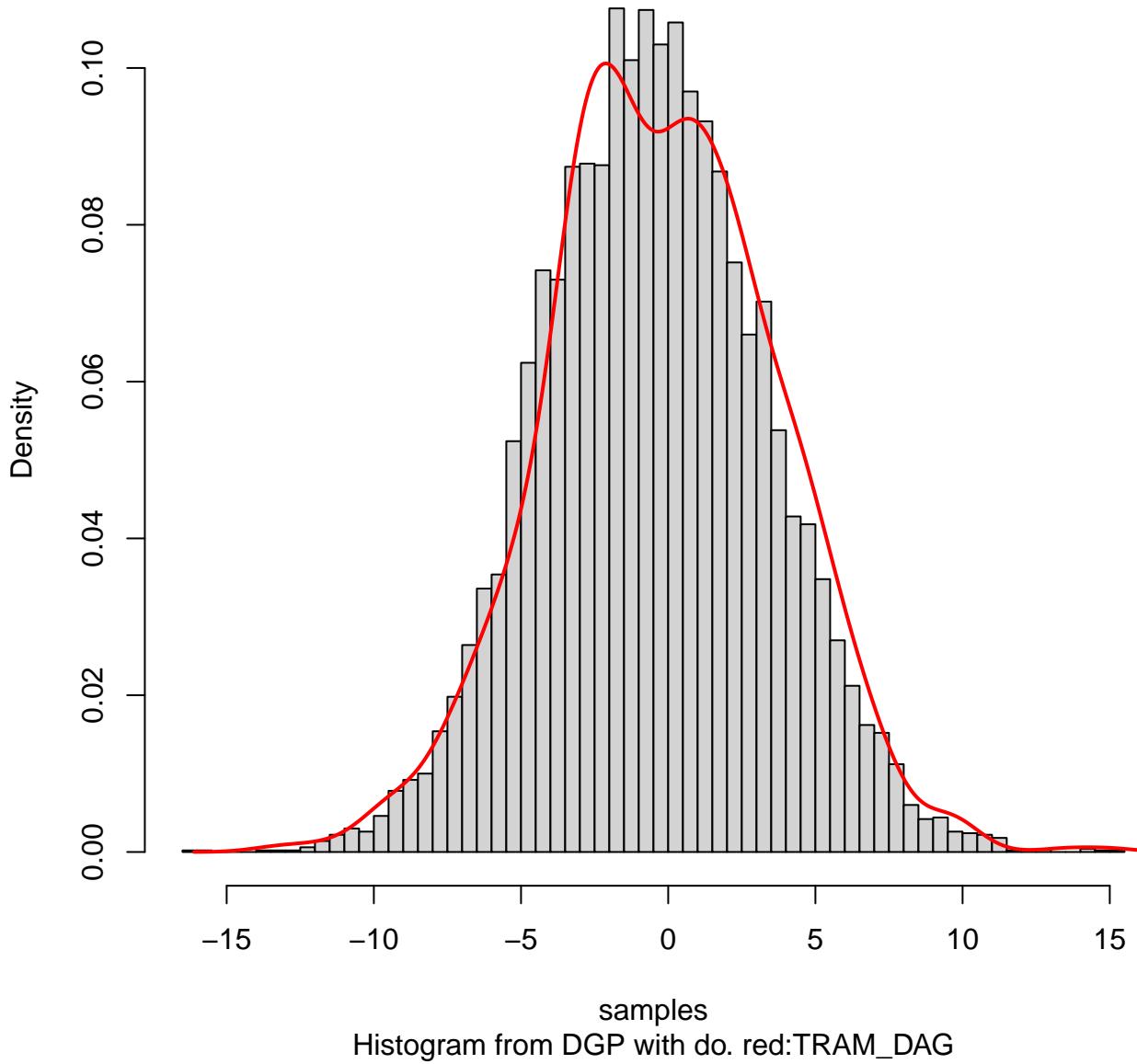
**Do(X1=0.2) X2**



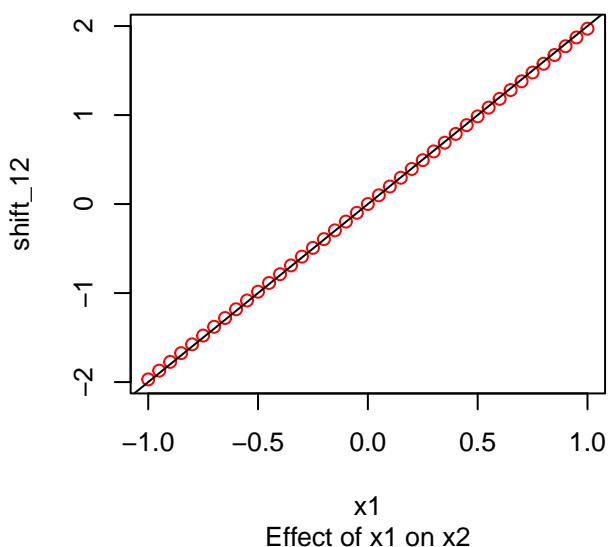
$X2 | Do(X1=0.2)$



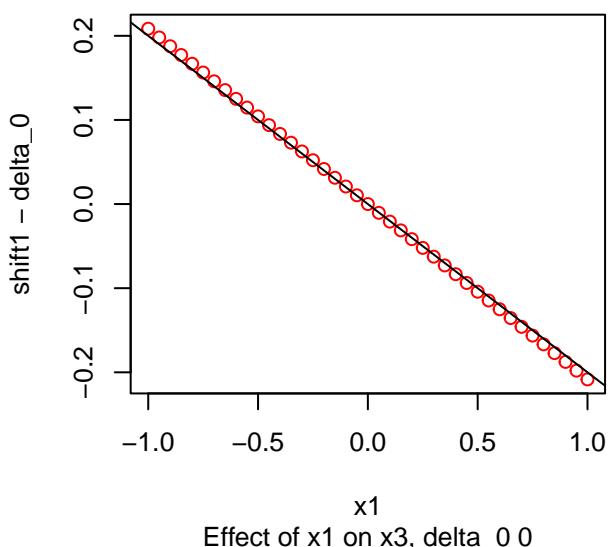
$X3 | \text{Do}(X1=0.2)$



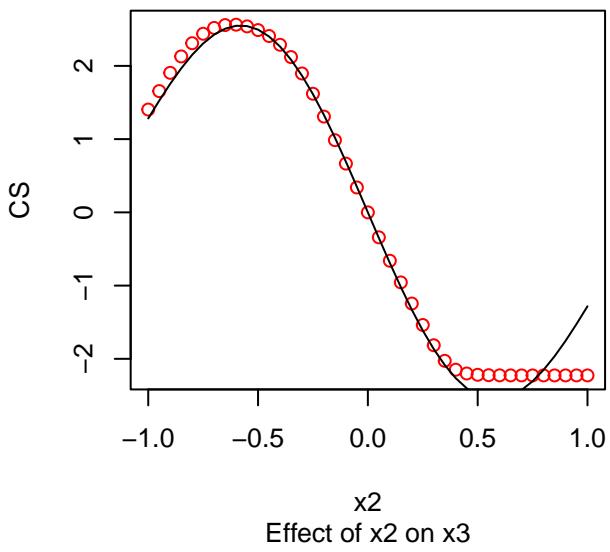
**LS-Term (black DGP, red Ours)**

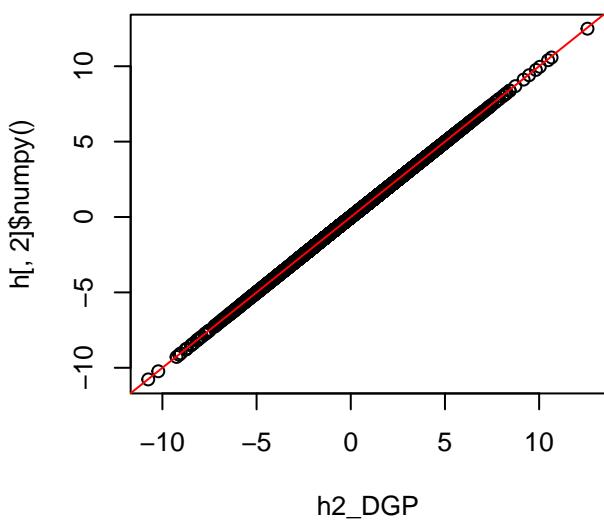
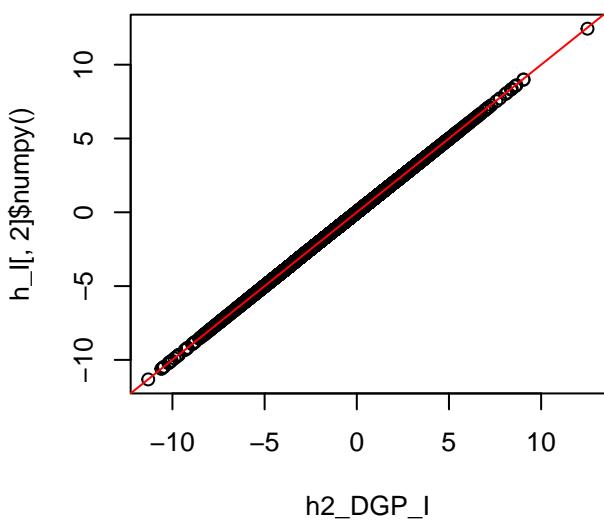
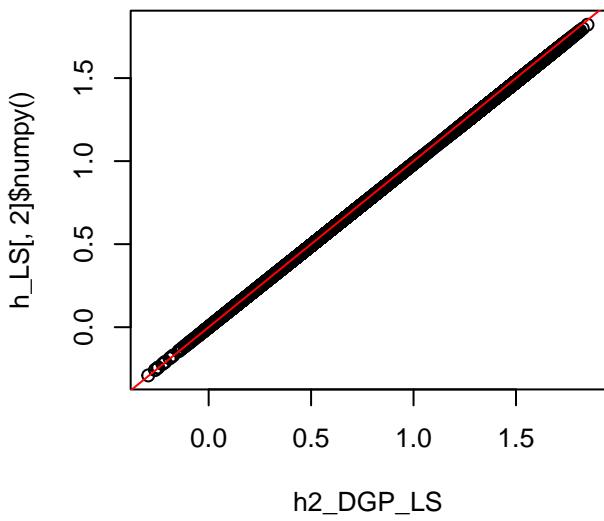
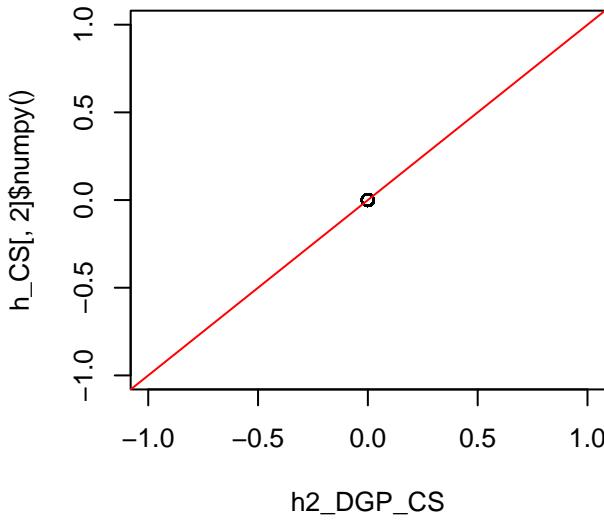


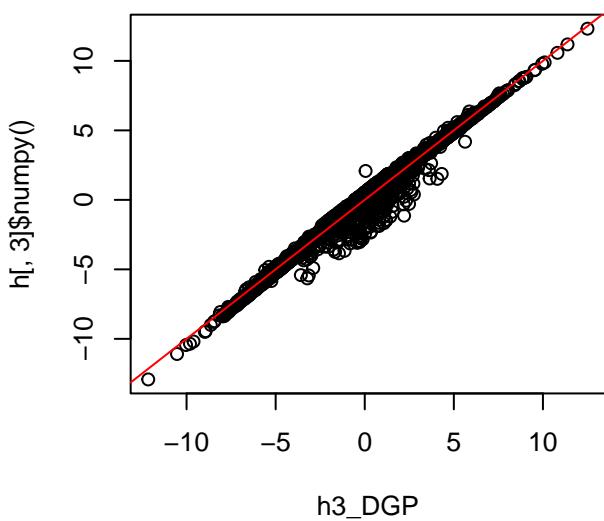
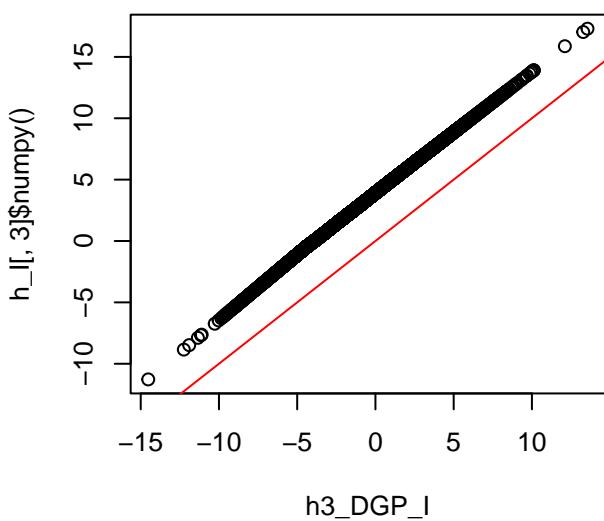
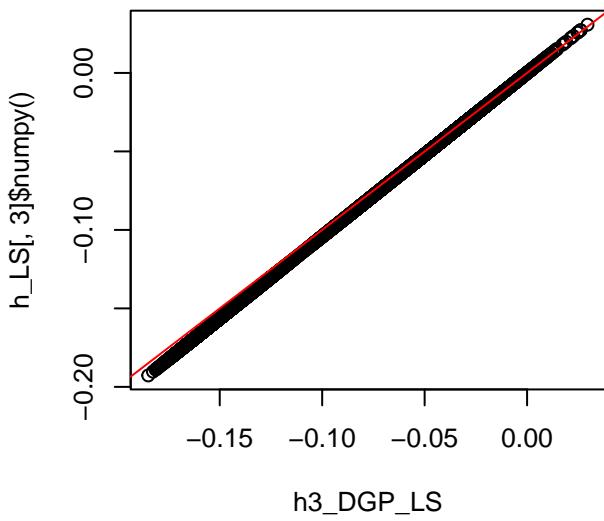
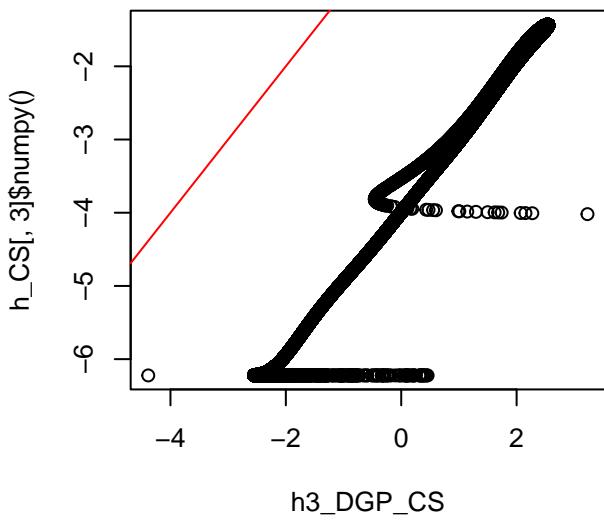
**LS-Term (black DGP, red Ours)**



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



**h2****h2\_I****h2\_LS****h2\_CS**

**h3****h3\_DGP****h3\_I****h3\_DGP\_I****h3\_LS****h3\_DGP\_LS****h3\_CS****h3\_DGP\_CS**