CAUSAL ADJUSTMENT METHODS USING THE PROPENSITY SCORE

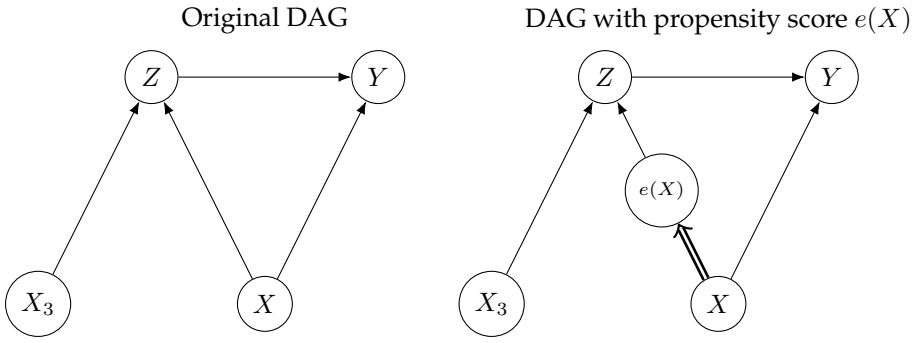
The most common causal inference problem is to estimate the causal effect of treatment Z on outcome Y in the presence of confounders X. The causal effect is defined via the expectation under the experimental measure ε.

$$\mu(z) = \mathbb{E}^{\varepsilon}_{Y|Z}[Y|Z=z] \qquad z \in \mathcal{Z}.$$

The propensity score for binary exposure, denoted $e(X)$, is defined via the observational conditional model for Z given confounders X, as

$$e(x) = f^{\mathcal{O}}_{Z|X}(1|x) = \Pr^{\mathcal{O}}_{Z|X}[Z=1|X=x]$$

with $e(X)$ being the corresponding random variable. The propensity score is a *balancing score*, that is, we have that $Z \perp\!\!\!\perp X \mid e(X)$. In the simulations below, we set $X = (X_1, X_2)$, but also introduce a third predictor X_3 which is not a confounder, but instead is purely a cause (or predictor) of Z.



Consider the following example where we generate according to the parametric model

- $(X_1, X_2, X_3)^\top \sim \text{Normal}_3(\mu, \Sigma)$ with

$$\mu = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.25 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 \end{bmatrix} \begin{bmatrix} 1.0 & 0.9 & 0.0 \\ 0.9 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.25 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 \end{bmatrix}$$

- $Z|X = x \sim Bernoulli(e(x;\alpha))$, with

$$e(x;\alpha) = \frac{\exp\{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2\}}{1 + \exp\{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2\}}$$

  with $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top = (11.5, -0.2, 2.7, 2, 2)^\top$.

- $Y|X = x, Z = z \sim \text{Normal}(\mu(x, z; \beta, \psi), \sigma^2)$, with

$$\mu(x,z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + z(\psi_0 + \psi_1 x_1 + \psi_2 x_2 + \psi_3 x_1 x_2) = \mu_0(x;\beta) + z\mu_1(x;\psi)$$

  say, with $\sigma^2 = 0.1^2$, and

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top = (10, -2.0, 1.2, 0.6)^\top \qquad \psi = (\psi_0, \psi_1, \psi_2, \psi_3)^\top = (1, 1, 1, 1)^\top$$

  In this model, X_1 and X_2 are confounders, and X_3 is a predictor of Z alone (and therefore is an *instrument*); X_3 is independent of (X_1, X_2).

```
#Data simulation
library(mvnfast);library(xtable)
set.seed(23987)
n<-5000
D<-diag(c(0.25,0.5,0.75))
Mu<-c(1,-2,-1)
Sigma<-D %*% (matrix(c(1,0.9,0.0,0.9,1,0.0,0,0,1),3,3) %*% D)
X<-rmvn(n,mu=Mu,Sigma)
```

```
al<-c(11.5,-0.2,2.7,2,2)
Xal<-cbind(1,X[,1],X[,2],X[,3],X[,1]*X[,2])
expit<-function(x){return(1/(1+exp(-x)))}
ps.true<-expit(Xal %*% al)
Z<-rbinom(n,1,ps.true)
be<-c(10,-2.0,1.2,0.6)
Xb<-cbind(1,X[,1],X[,2],X[,1]*X[,2])
mu0<- Xb %*% be
psi<-c(1,1,1,1)
Xp<-cbind(1,X[,1],X[,2],X[,1]*X[,2])
mu1<- (Xp %*% psi)
eta<-mu0 + Z * mu1 ; sig<-0.1
Y<-rnorm(n,eta,sig)
X1<-X[,1];X2<-X[,2];X3<-X[,3]
true.vals<-c(be,psi)[c(1,2,3,5,4,6,7,8)]
eX<-as.vector(Z-ps.true)
```

The average treatment effect (ATE) can be computed directly in terms of the model parameters:

$$\mu(z) = \mathbb{E}^{\varepsilon}_{Y|Z}[Y|Z=z] \equiv \mathbb{E}^{\mathcal{O}}_{X}\left[\mathbb{E}^{\mathcal{O}}_{Y|X,Z}[Y|X,Z=z]\right] = \mathbb{E}^{\mathcal{O}}_{X}\left[\mu(X,z)\right]$$

so in this case

$$\begin{aligned}\mu(z) &= \mathbb{E}^{\mathcal{O}}_{X}\left[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + z(\psi_0 + \psi_1 X_1 + \psi_2 X_2 + \psi_3 X_1 X_2)\right] \\ &= \beta_0 + \beta_1 \mathbb{E}[X_1] + \beta_2 \mathbb{E}[X_2] + \beta_3 \mathbb{E}[X_1 X_2] + z(\psi_0 + \psi_1 \mathbb{E}[X_1] + \psi_2 \mathbb{E}[X_2] + \psi_3 \mathbb{E}[X_1 X_2])\end{aligned}$$

Here we have that

$$\mathbb{E}[XX^\top] = \begin{bmatrix} \mathbb{E}[X_1^2] & \mathbb{E}[X_1X_2] & \mathbb{E}[X_1X_3] \\ \mathbb{E}[X_1X_2] & \mathbb{E}[X_2^2] & \mathbb{E}[X_2X_3] \\ \mathbb{E}[X_1X_3] & \mathbb{E}[X_2X_2] & \mathbb{E}[X_3^2] \end{bmatrix} = \Sigma + \mu\mu^\top = \begin{bmatrix} 0.0625 & 0.1125 & 0 \\ 0.1125 & 0.25 & 0 \\ 0 & 0 & 0.5625 \end{bmatrix} + \begin{bmatrix} 1 & -2 & -1 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{bmatrix}$$

which yields

```
(X.M<-Sigma + Mu %*% t(Mu))

##         [,1]    [,2]    [,3]
## [1,]  1.0625 -1.8875 -1.0000
## [2,] -1.8875  4.2500  2.0000
## [3,] -1.0000  2.0000  1.5625
```

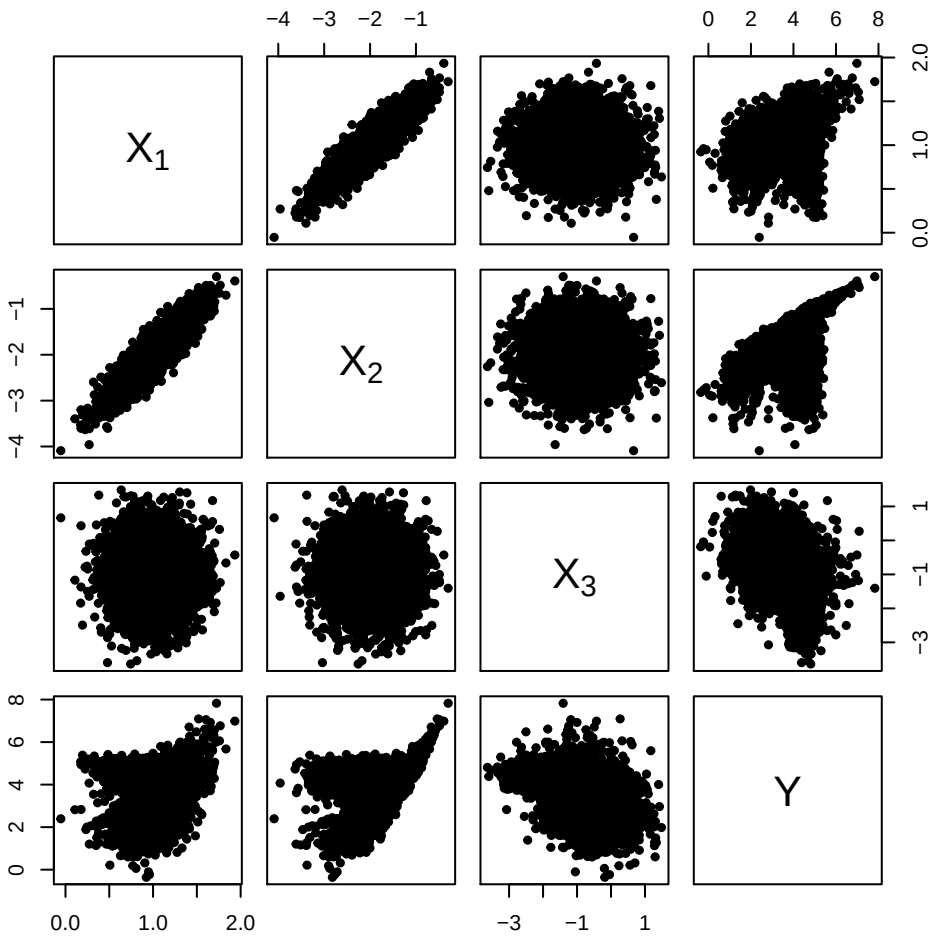
so therefore $\mu(0)$ and $\mu(1)$ are

```
(mu0<-sum(be*c(1,Mu[1],Mu[2],X.M[1,2])))

## [1] 4.4675

(mu1<-mu0 + sum(psi*c(1,Mu[1],Mu[2],X.M[1,2])))

## [1] 2.58
```

and hence the 'causal effect' is

$$\mu(1) - \mu(0) = 2.58 - 4.4675 = -1.8875.$$

```
par(mar=c(3,3,3,3))
pairs(cbind(X,Y),pch=19,cex=0.75,
      labels=c(expression(X[1]),expression(X[2]),expression(X[3]),expression(Y)))
```



A correctly specified linear regression model recovers the correct coefficients. The estimated ψ values are given in the following output:

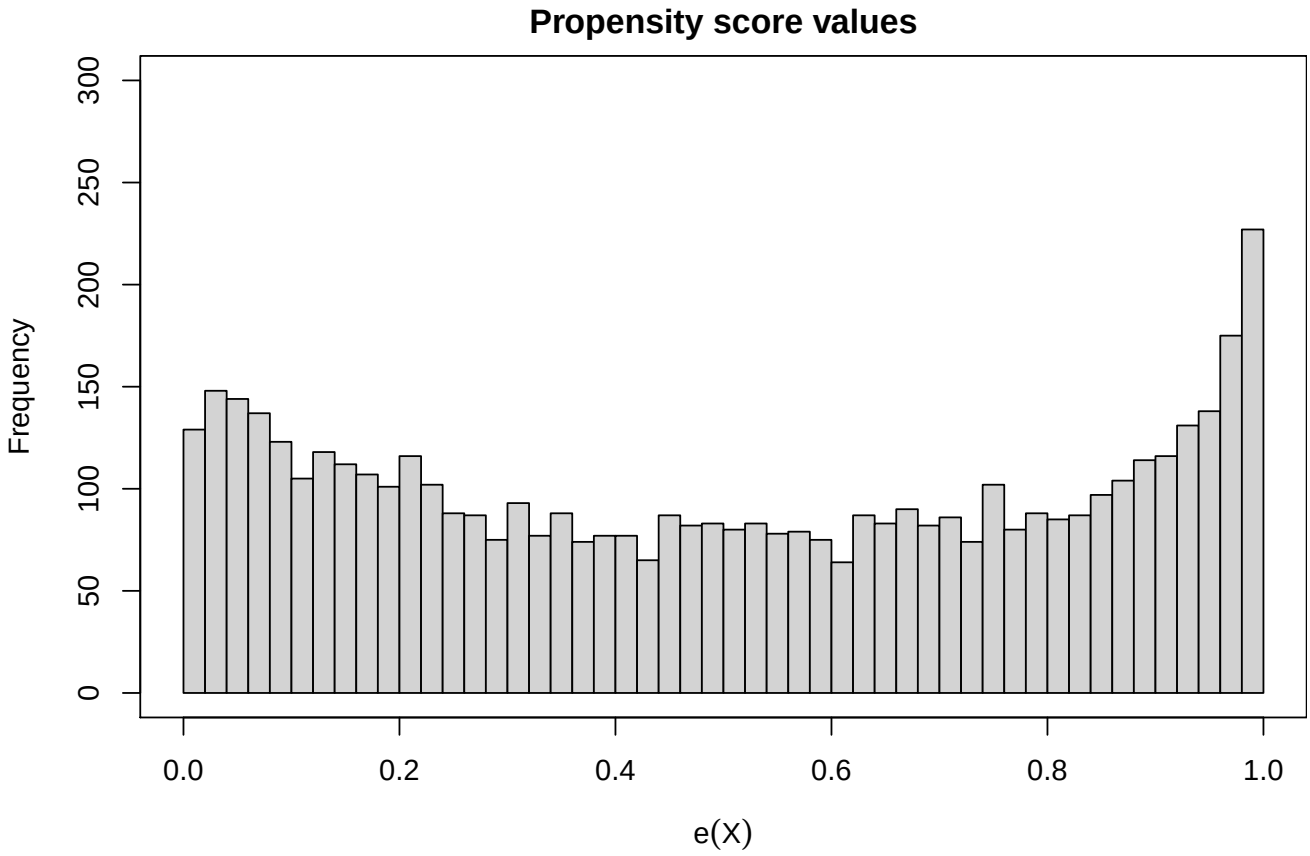
```
round(coef(summary(lm(Y~X1+X2+X1:X2+Z+Z:X1+Z:X2+Z:X1:X2)))[c(4,6:8),],6)

##          Estimate Std. Error  t value Pr(>|t|)
## Z        1.036320   0.066489 15.58638        0
## X1:Z     0.988930   0.044025 22.46274        0
## X2:Z     1.006031   0.021809 46.12957        0
## X1:X2:Z 1.006289   0.016679 60.33373        0
```

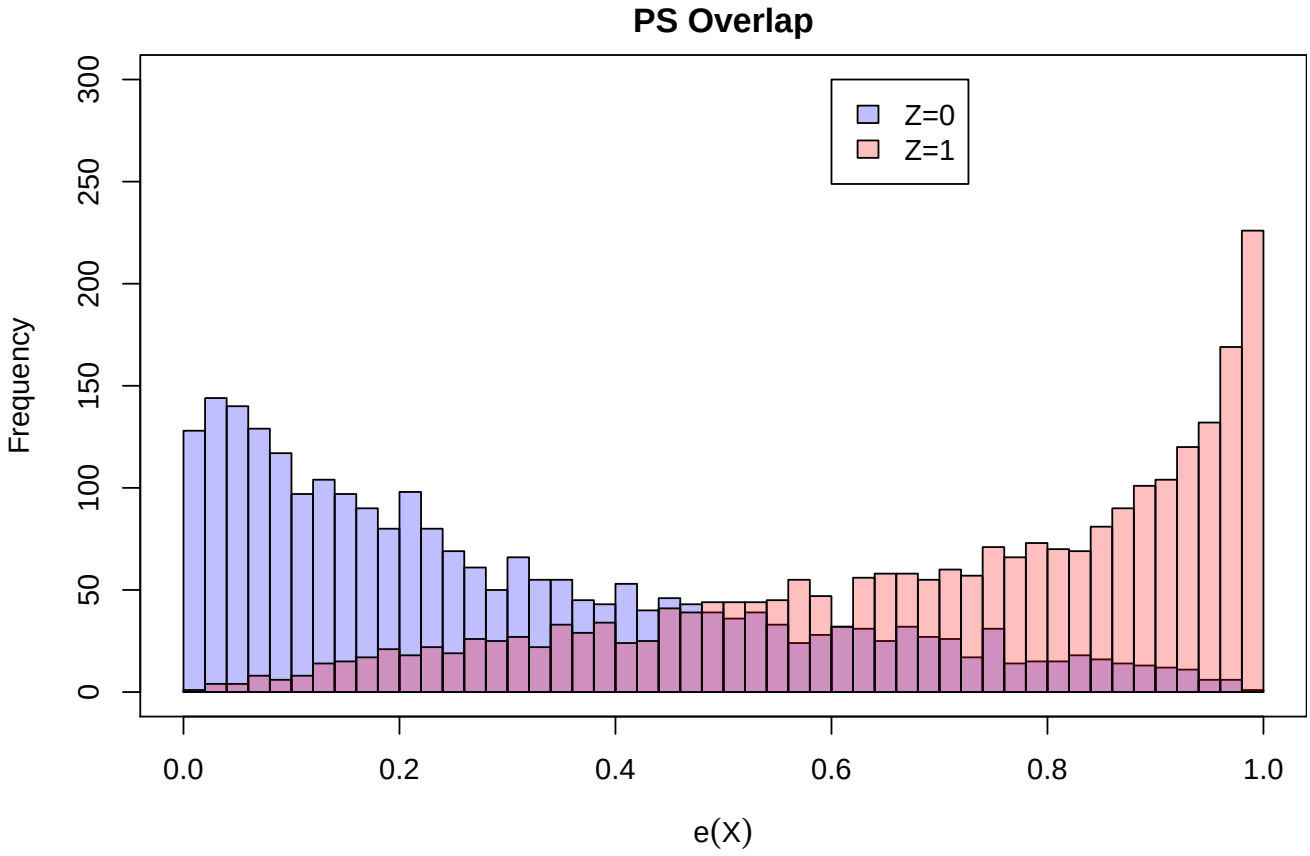
We now study the properties of the propensity score in this case. We need to inspect

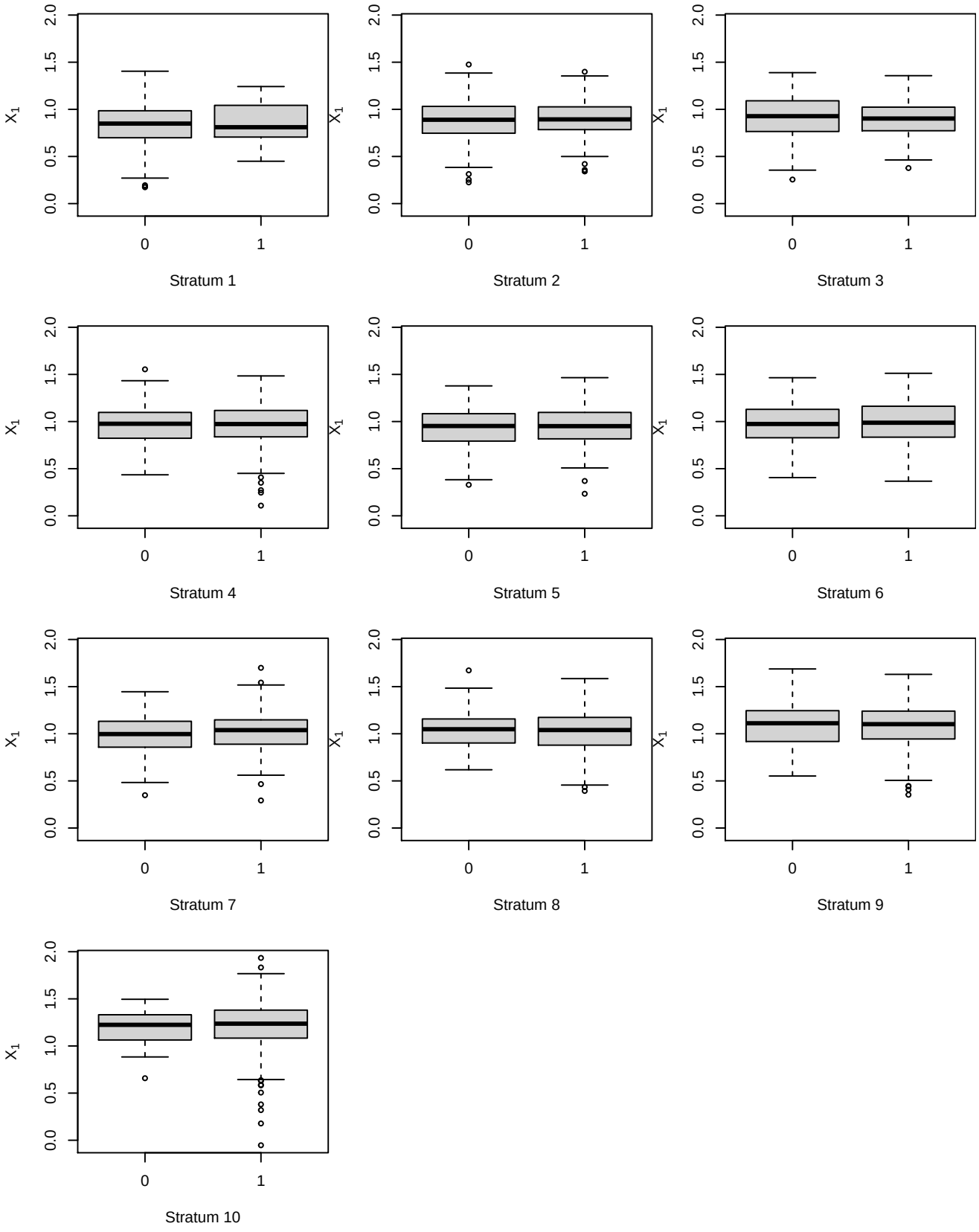
- propensity score **overlap**: assess whether the propensity score distributions for the two treatment groups are sufficiently overlapping to allow comparison within propensity score strata.
- **balance:** assess whether, within strata of the propensity score the distribution of X_1 and X_2 is the same within $Z = 0$ and $Z = 1$ groups.

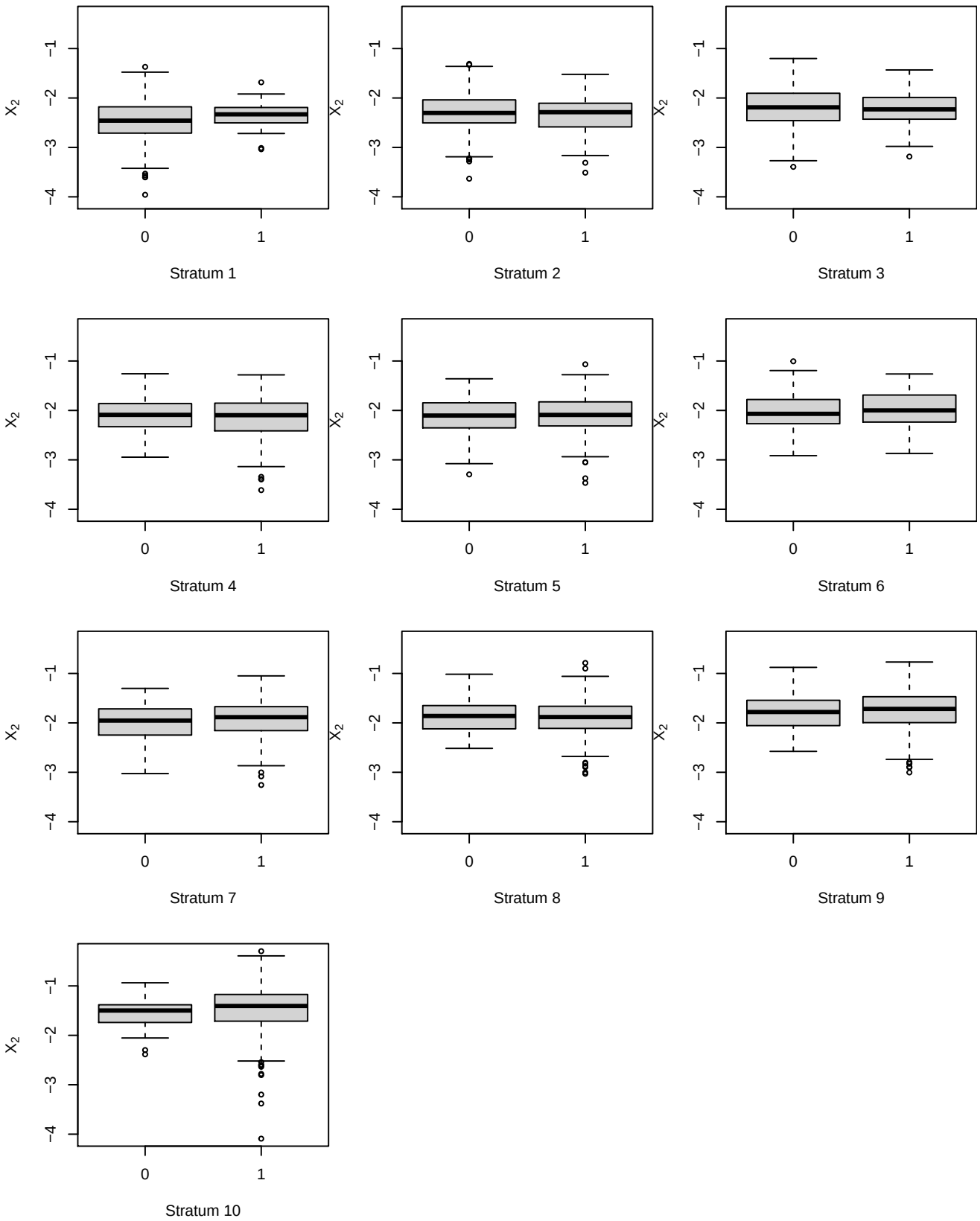
Here we define the strata by taking the partition based on intervals of width 0.1 of the propensity score distribution. There is a reasonable spread of data across these strata.

```
ps.cat<-cut(ps.true,seq(0,1,by=0.1),labels=FALSE);table(Z,ps.cat)

##    ps.cat
## Z     1   2   3   4   5   6   7   8   9  10
##   0 658 468 358 264 221 160 147 103  76  36
##   1  23  75 110 145 173 235 259 327 411 751
```


Propensity score values
Frequency
0
50
100
150
200
250
300
0.0
0.2
0.4
0.6
0.8
1.0
e(X)


PS Overlap
Z=0
Z=1
Frequency
0
50
100
150
200
250
300
0.0
0.2
0.4
0.6
0.8
1.0
e(X)


X1
0.0
0.5
1.0
1.5
2.0
0
1
Stratum 1
Stratum 2
Stratum 3
Stratum 4
Stratum 5
Stratum 6
Stratum 7
Stratum 8
Stratum 9
Stratum 10

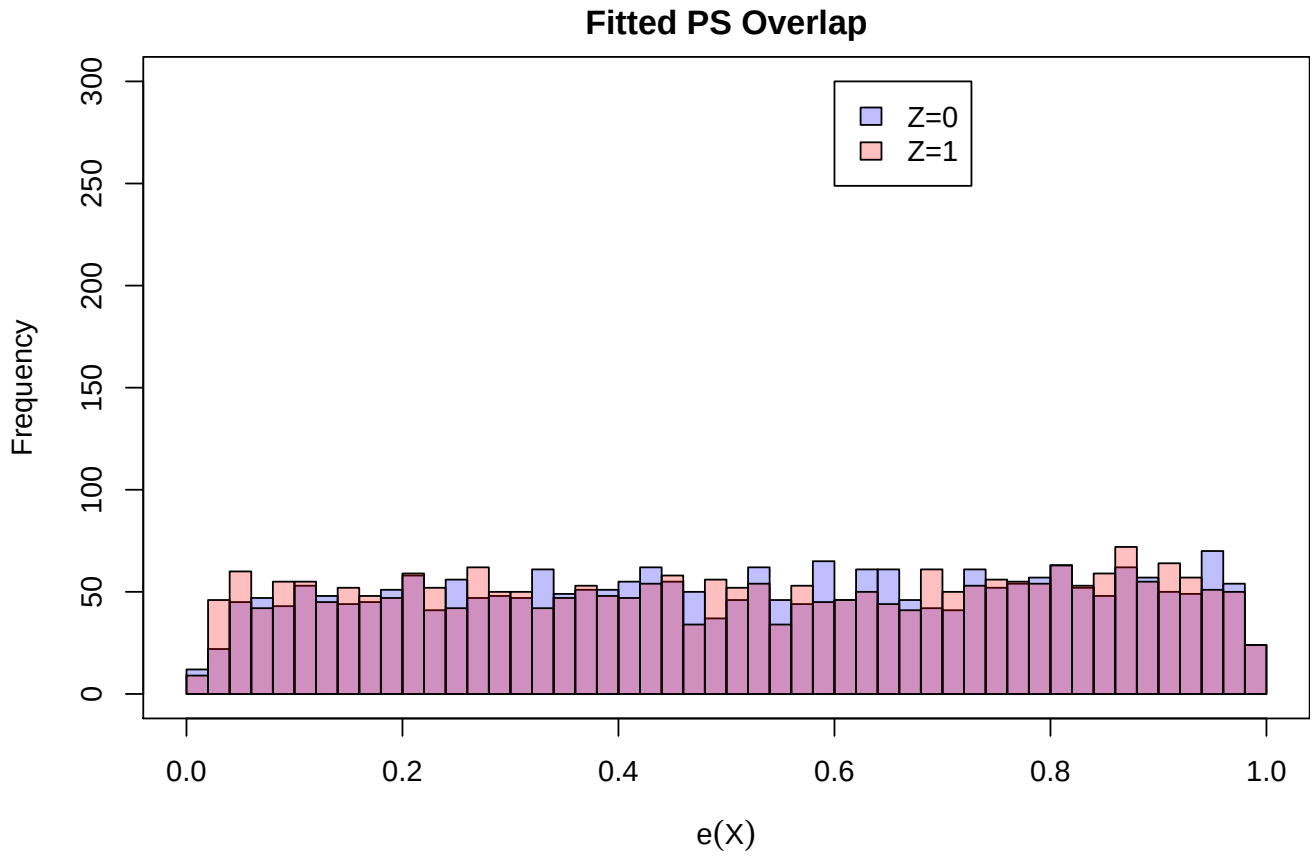


There is reasonable propensity score overlap for the two treatment groups.

If we instead consider fitted values of the propensity score, obtained using logistic regression, the overlap is

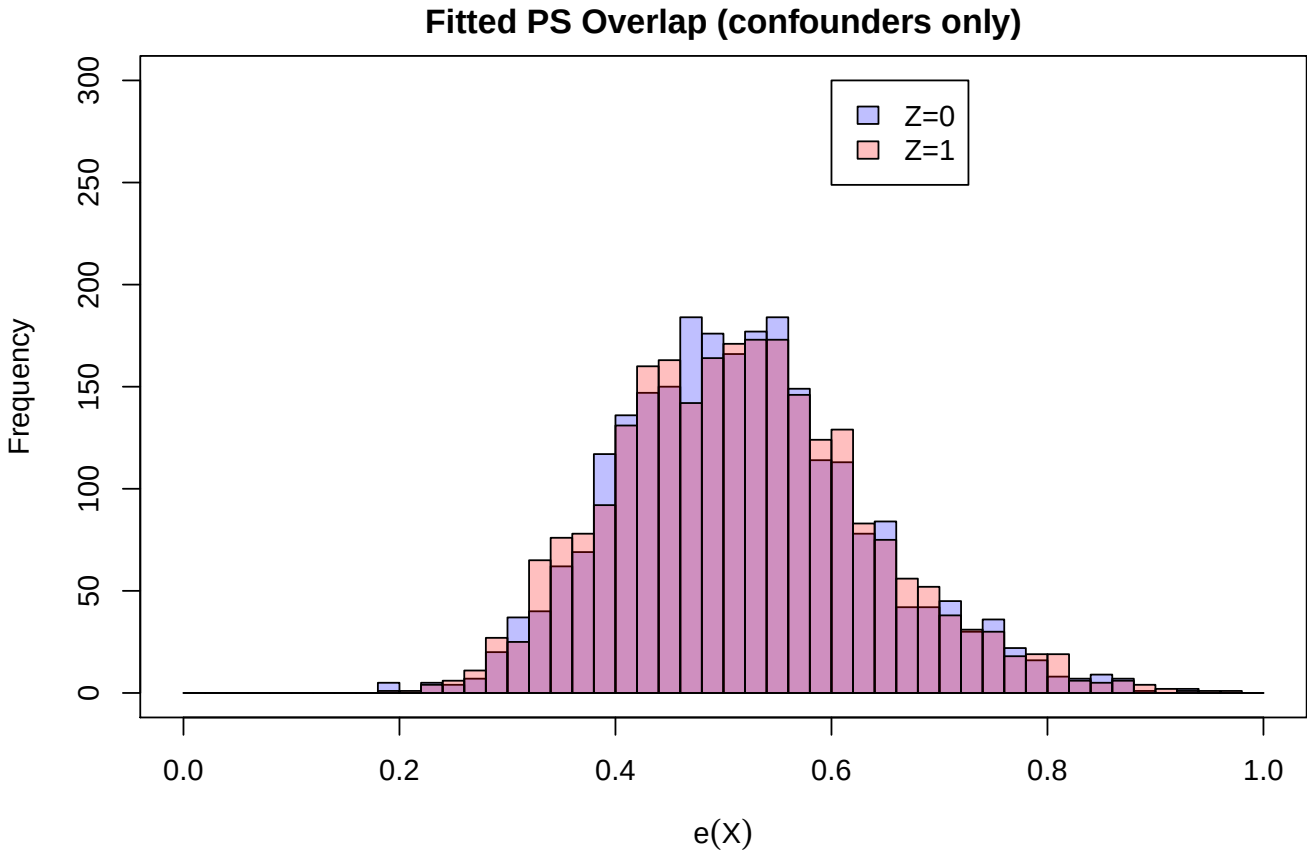
essentially unaffected.

```
ps.fit1<-fitted(fit1<-glm(Z~X1+X2+X3+X1:X2,family=binomial))
ps.cat1<-cut(ps.fit1,seq(0,1,by=0.1),labels=FALSE)
round(coef(summary(fit1)),6)

##              Estimate Std. Error   z value Pr(>|z|)
## (Intercept)  7.005232   0.781352  8.965522 0.000000
## X1          -0.176789   0.542310 -0.325992 0.744430
## X2           0.822602   0.263859  3.117578 0.001823
## X3           1.958894   0.062220 31.483536 0.000000
## X1:X2        1.660596   0.216255  7.678872 0.000000
```



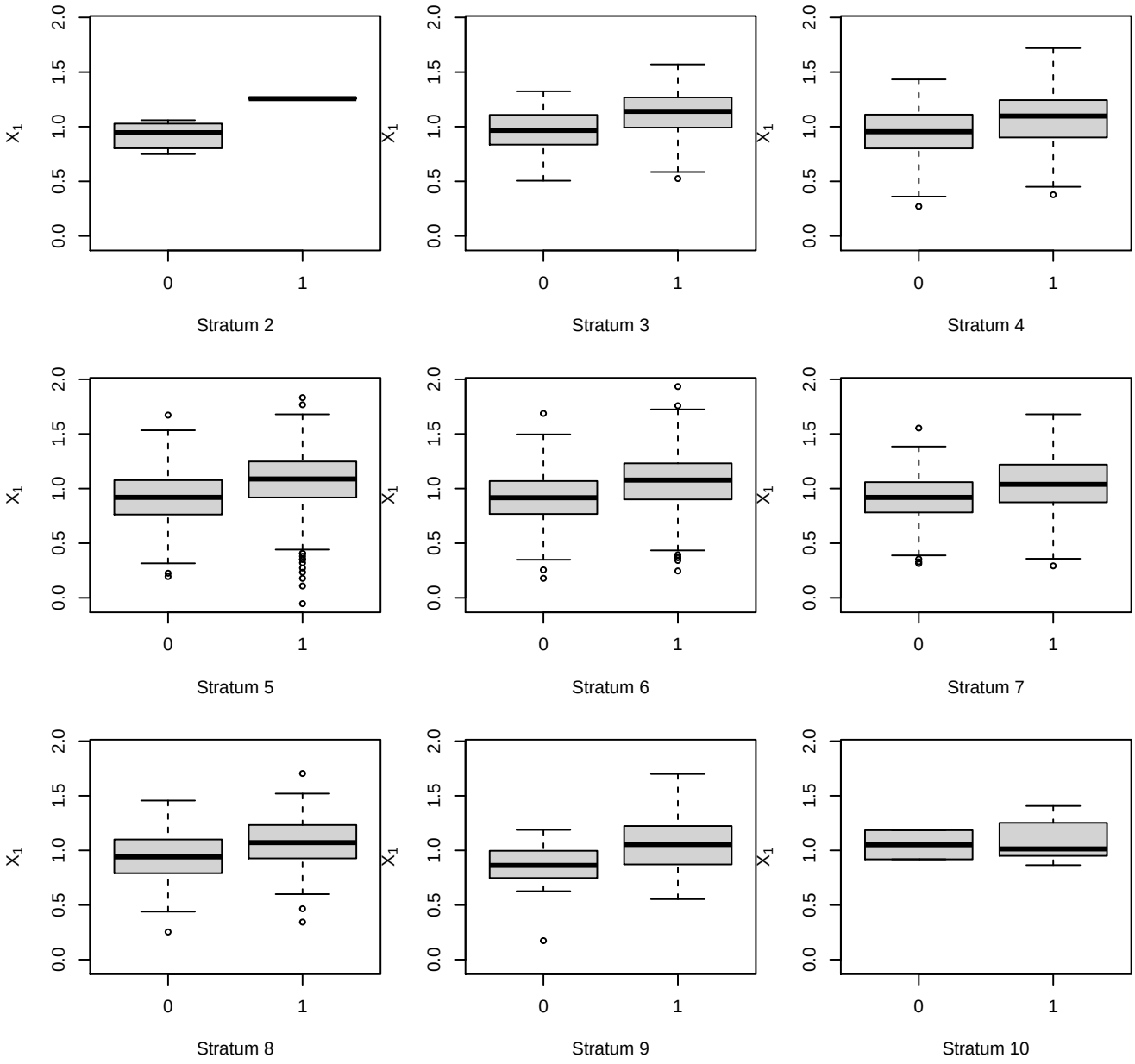
However, theory suggests that we can **omit** X_3 from the propensity score model as X_3 is not a confounder. This changes the fitted propensity score distribution considerably.

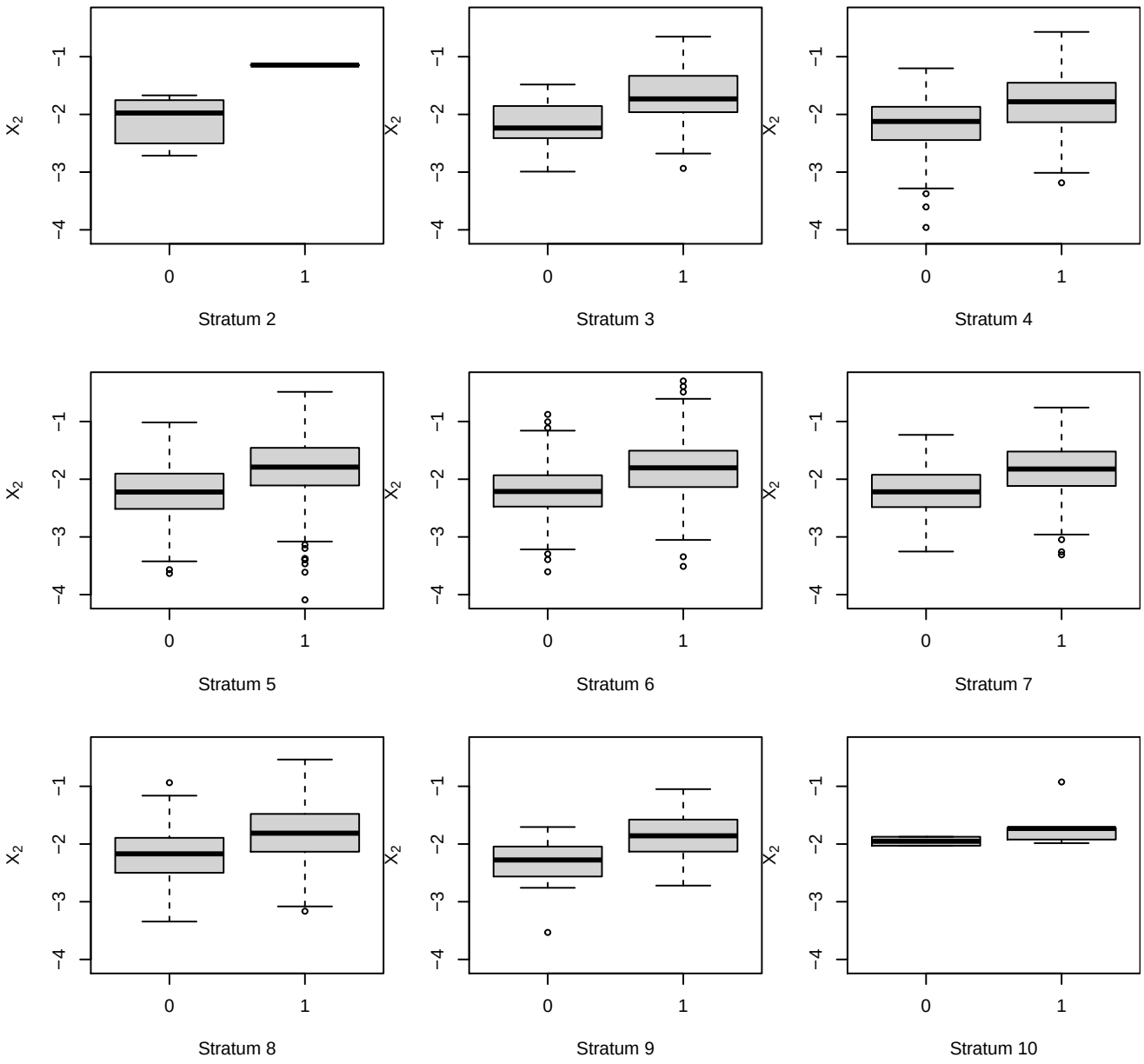
```
ps.fit2<-fitted(fit2<-glm(Z~X1+X2+X1:X2,family=binomial))
round(coef(summary(fit2)),6)

##              Estimate Std. Error   z value Pr(>|z|)
## (Intercept)  4.111006   0.653285  6.292822 0.000000
## X1          -0.313443   0.457468 -0.685170 0.493237
## X2           0.700421   0.222224  3.151867 0.001622
## X1:X2        1.228501   0.181193  6.780056 0.000000
```



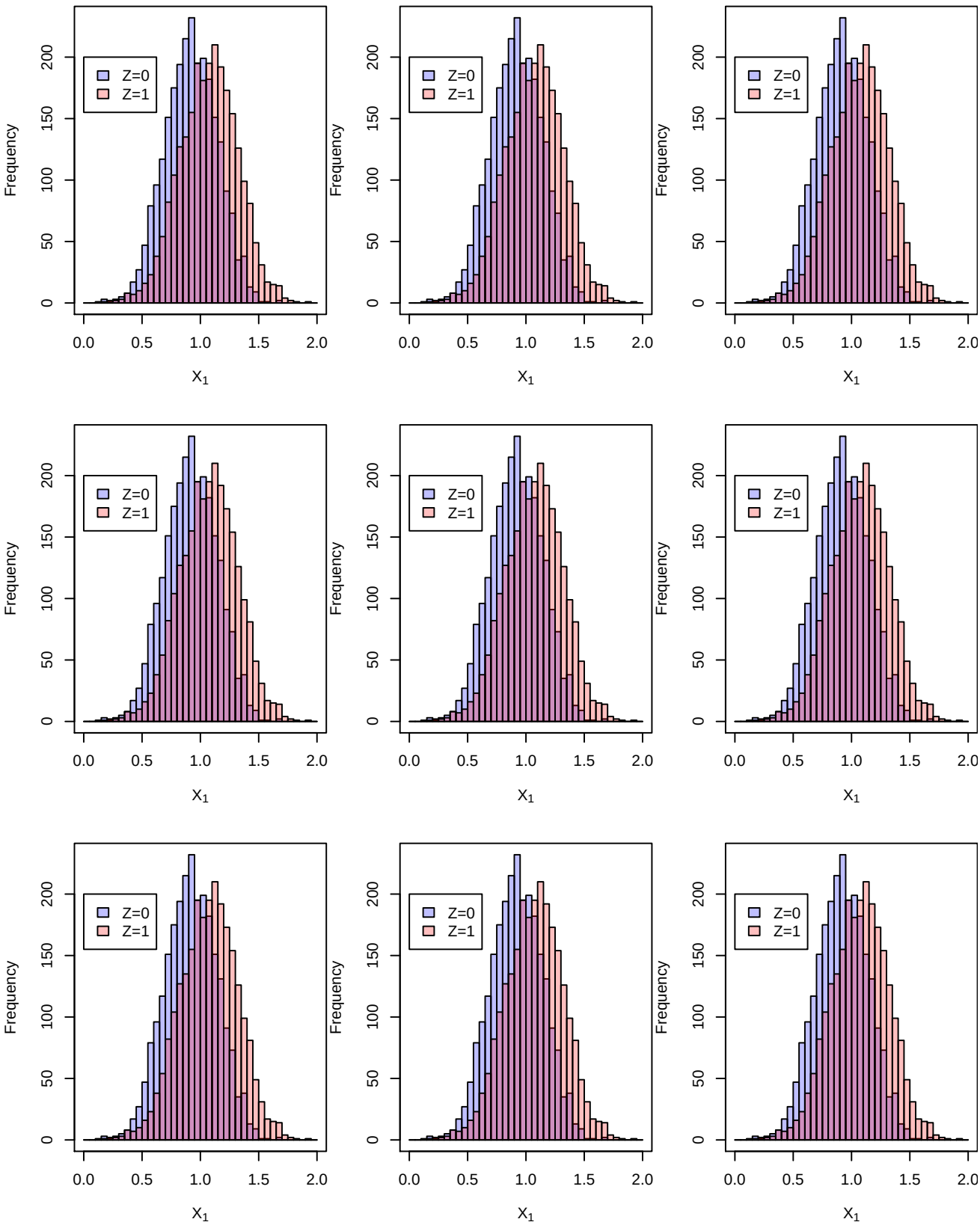
The overlap is much better. Balance is still largely induced in the confounder distributions, although the strata near the ends of the range are almost empty.

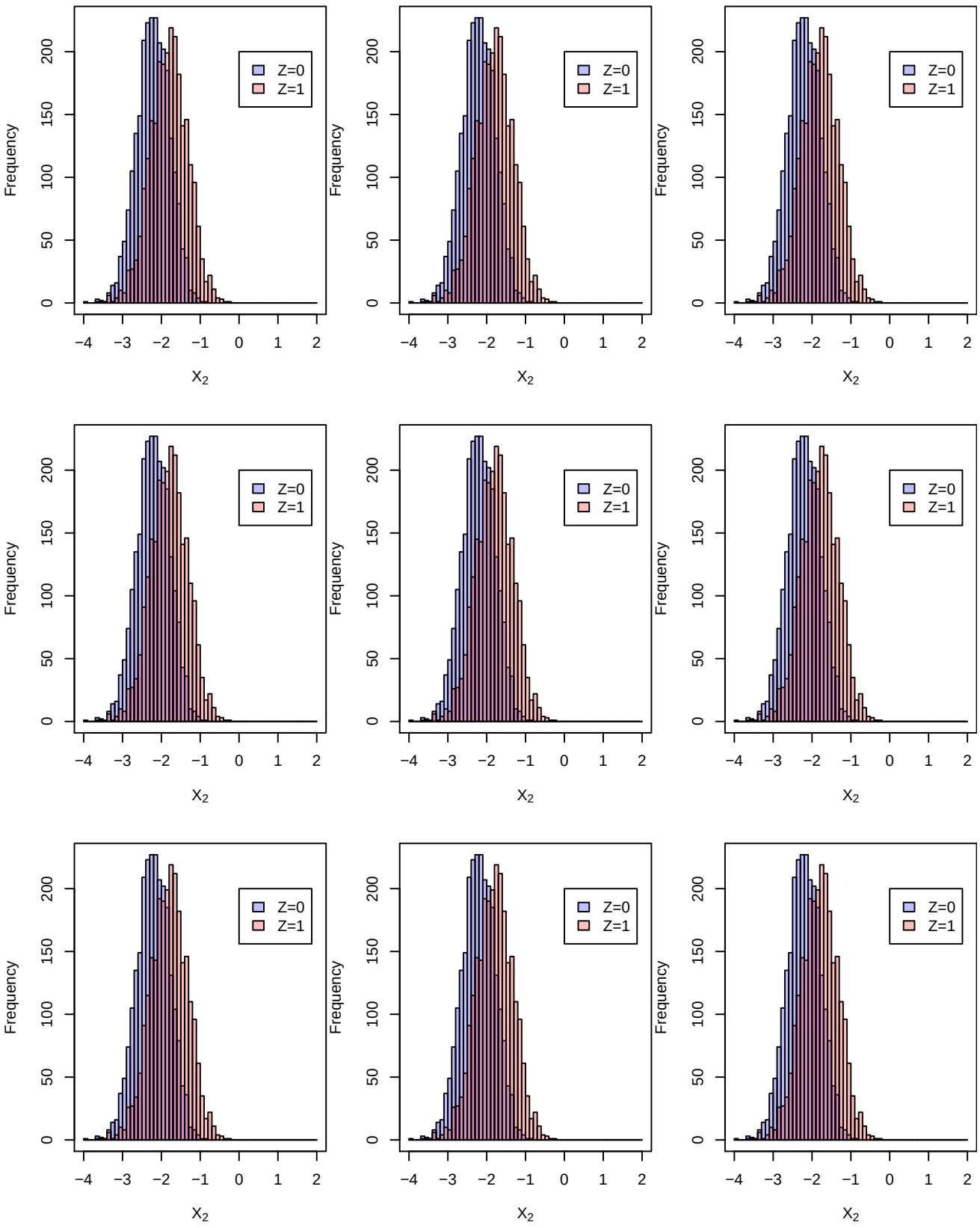
```
##    ps.cat2
## Z     2   3   4   5   6   7   8   9  10
##   0   5  36 325 793 790 359 149  32   2
##   1   1  49 336 760 787 395 136  40   5
```


X1
0.0
0.5
1.0
1.5
2.0
0
1
Stratum 2
Stratum 3
Stratum 4
Stratum 5
Stratum 6
Stratum 7
Stratum 8
Stratum 9
Stratum 10



Evidently, the overlap is reasonably good, but the boxplots imply a slight lack of balance. We can investigate further using histograms.


Z=0
Z=1
Frequency
0
50
100
150
200
0.0
0.5
1.0
1.5
2.0
X1



The apparent lack of balance is perhaps due to the fairly coarse propensity score categories.

The final question is whether we can recover good estimates using statistical approaches based on the propensity

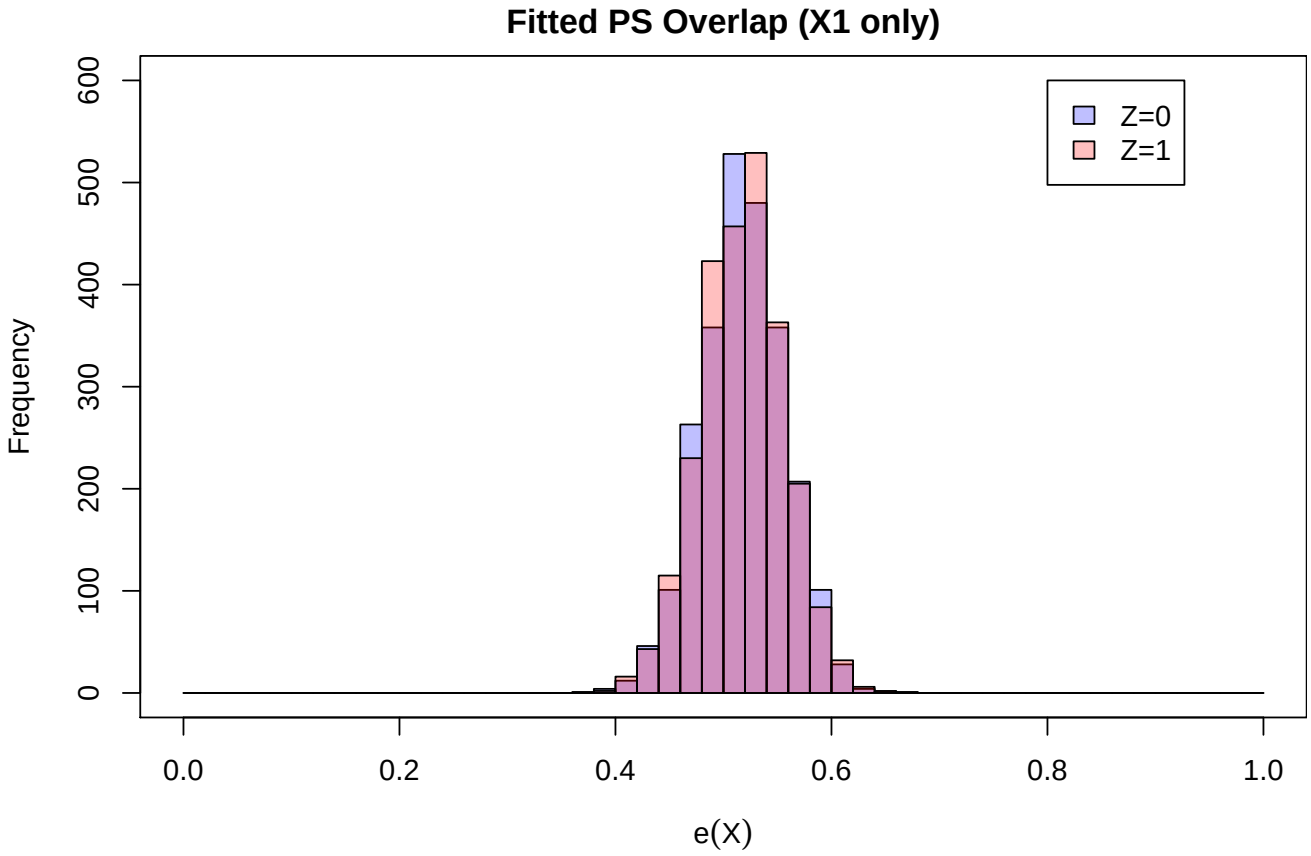
score. We first attempt stratification-based estimation of the ATE.

```
#True ps
local.ate<-rep(0,10)
for(j in 1:10){
    local.ate[j]<-mean(Y[ps.cat==j & Z == 1]) - mean(Y[ps.cat==j & Z == 0])
}
w0<-as.numeric(table(ps.cat)/n)
ate0<-sum(local.ate*w0)
ate0

## [1] -1.84563

#PS fit
local.ate1<-rep(0,10)
for(j in 1:10){
    local.ate1[j]<-mean(Y[ps.cat1==j & Z == 1]) - mean(Y[ps.cat1==j & Z == 0])
}
w1<-as.numeric(table(ps.cat1)/n)
ate1<-sum(local.ate*w1)
ate1

## [1] -1.874289

#PS fit with confounders only
local.ate2<-rep(0,10)
for(j in 1:10){
    #Watch for empty strata
    if(sum(ps.cat2==j & Z==1)*sum(ps.cat2==j & Z==0) == 0){
        #Empty !
        local.ate2[j]<-0
    }else{
        local.ate2[j]<-mean(Y[ps.cat2==j & Z == 1]) - mean(Y[ps.cat2==j & Z == 0])
    }
}
w2<-as.numeric(tabulate(ps.cat2,10)/n)
ate2<-sum(local.ate*w2)
ate2

## [1] -1.924229
```

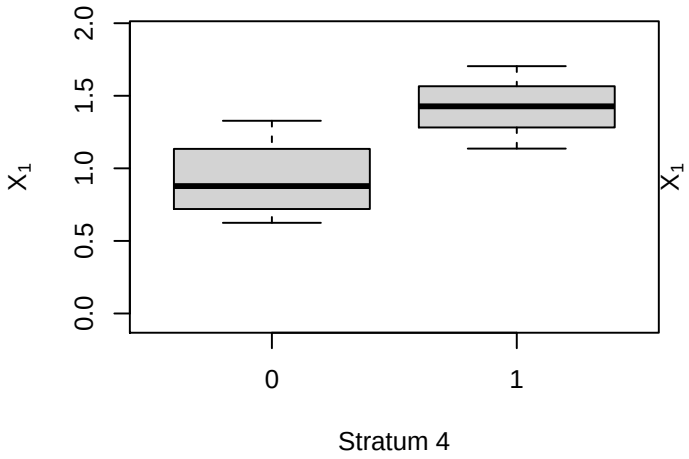
These raw estimates seem satisfactory and quite close to the true value of -1.8875, although further computation would be needed to assess standard errors.

For illustration, we demonstrate the results obtained when the propensity score model does not contain all the confounders.

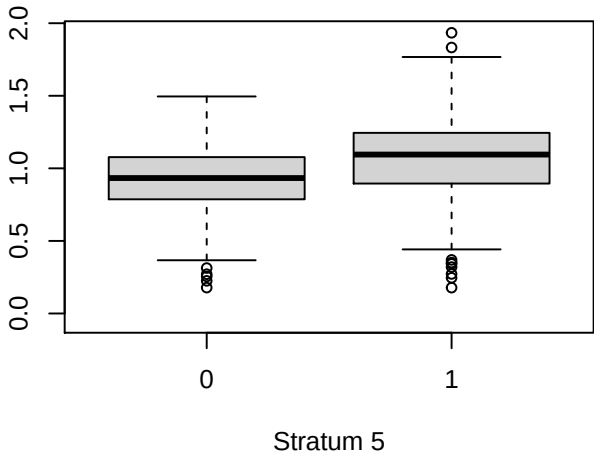
```
ps.fit3<-fitted(fit3<-glm(Z~X1,family=binomial))
round(coef(summary(fit3)),6)

##             Estimate Std. Error   z value Pr(>|z|)
## (Intercept) -0.546072   0.116299 -4.695422 0.000003
## X1           0.618932   0.113298  5.462883 0.000000
```



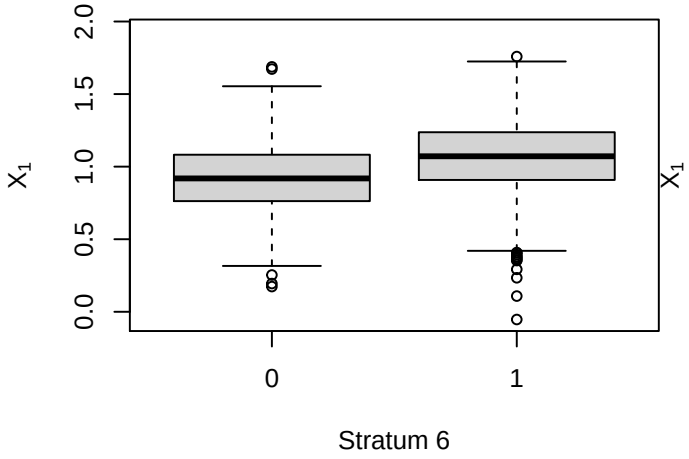

```
##    ps.cat3
## Z      4    5    6    7
##   0    4  780 1674   33
##   1    3  827 1638   41
```



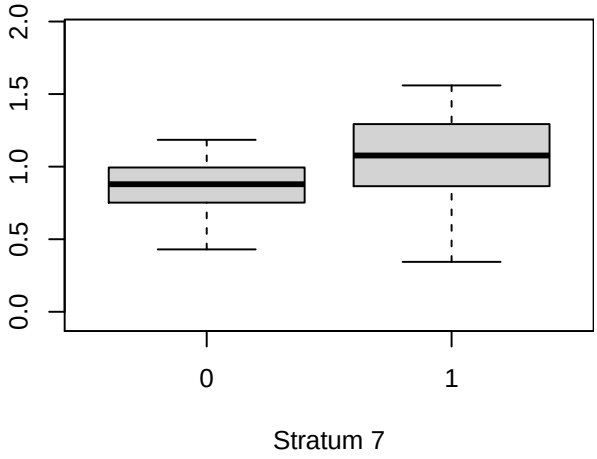
Stratum 4



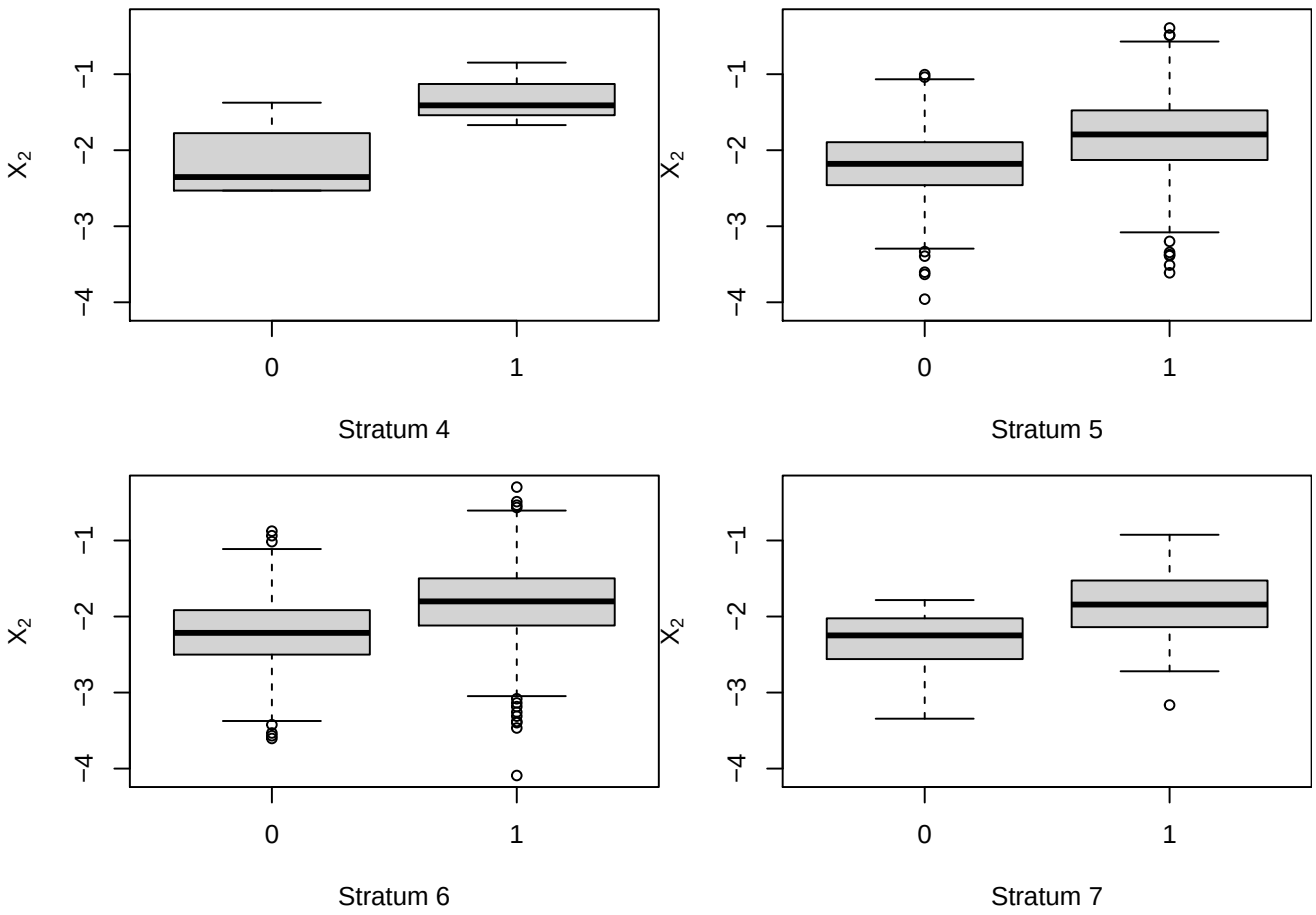
Stratum 5



Stratum 6



Stratum 7




```
#PS containing X1 only
local.ate3<-rep(0,10)
for(j in 1:10){
    #Watch for empty strata
    if(sum(ps.cat3==j & Z==1)*sum(ps.cat3==j & Z==0) == 0){
        #Empty !
        local.ate3[j]<-0
    }else{
        local.ate3[j]<-mean(Y[ps.cat3==j & Z == 1]) - mean(Y[ps.cat3==j & Z == 0])
    }
}
w3<-as.numeric(tabulate(ps.cat3,10)/n)
ate3<-sum(local.ate3*w3)
ate3

## [1] -1.305917
```

This point estimate is clearly far away from the true value of -1.8875.

We can now carry out a Monte Carlo study of the performance of the different estimation procedures: for 5000 replications with data sets of $n = 2000$, the following results display the results of four different models: using the propensity score regression model

$$\beta_0 + z(\psi_0 + \psi_1 x_1 + \psi_2 x_2 + \psi_3 x_1 x_2) + e(x)(\phi_0 + \phi_1 x_1 + \phi_2 x_2 + \phi_3 x_1 x_2)$$

we use for methods for constructing the propensity score

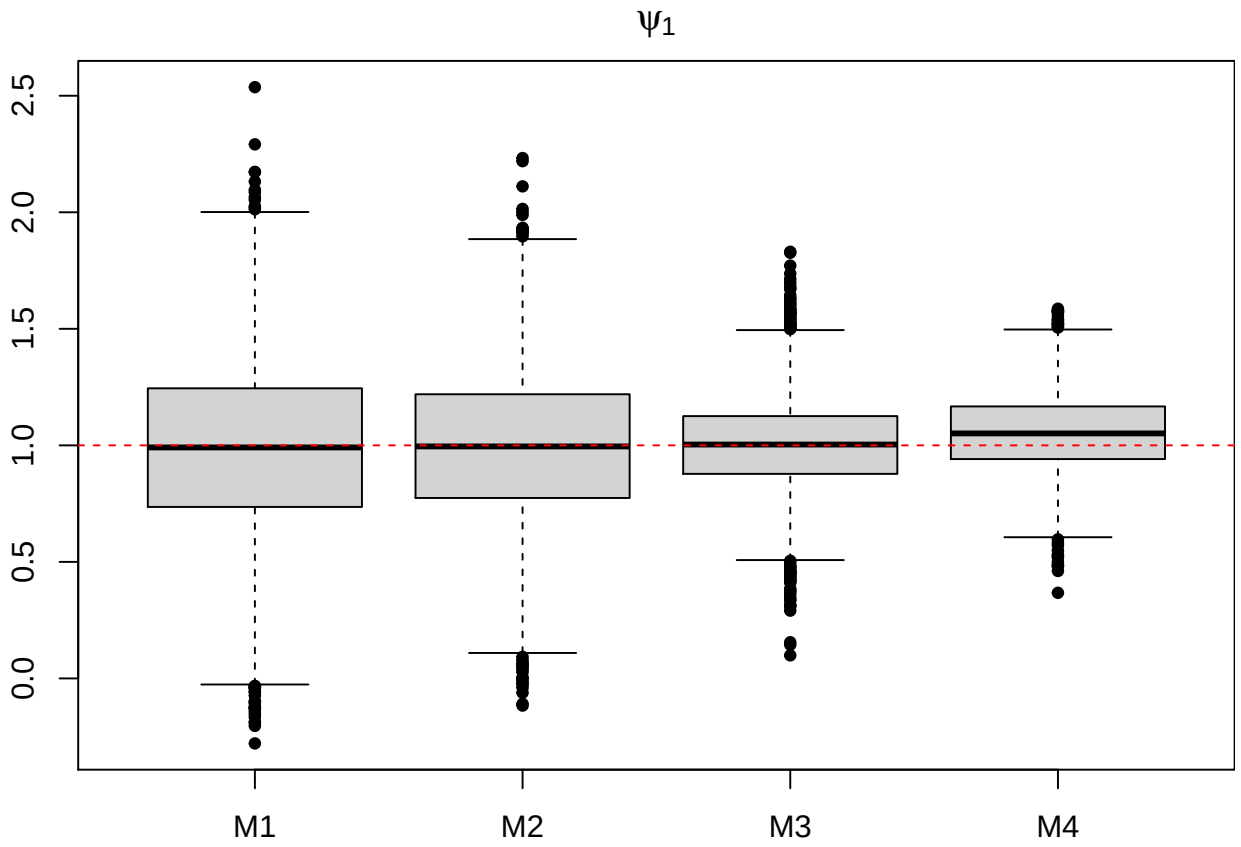
- M1: $e(x_1, x_2, x_3; \alpha)$ computed using the data generating α values;
- M2: $e(x_1, x_2, x_3; \widehat{\alpha})$ computed using fitted α values;
- M3: $e(x_1, x_2; \widehat{\alpha})$ computed using fitted α values omitting X_3;
- M4: $e(x_1; \widehat{\alpha})$ computed using fitted α values for a model only utilizing X_1.

The first three methods should give unbiased estimation, whereas the fourth should result in bias

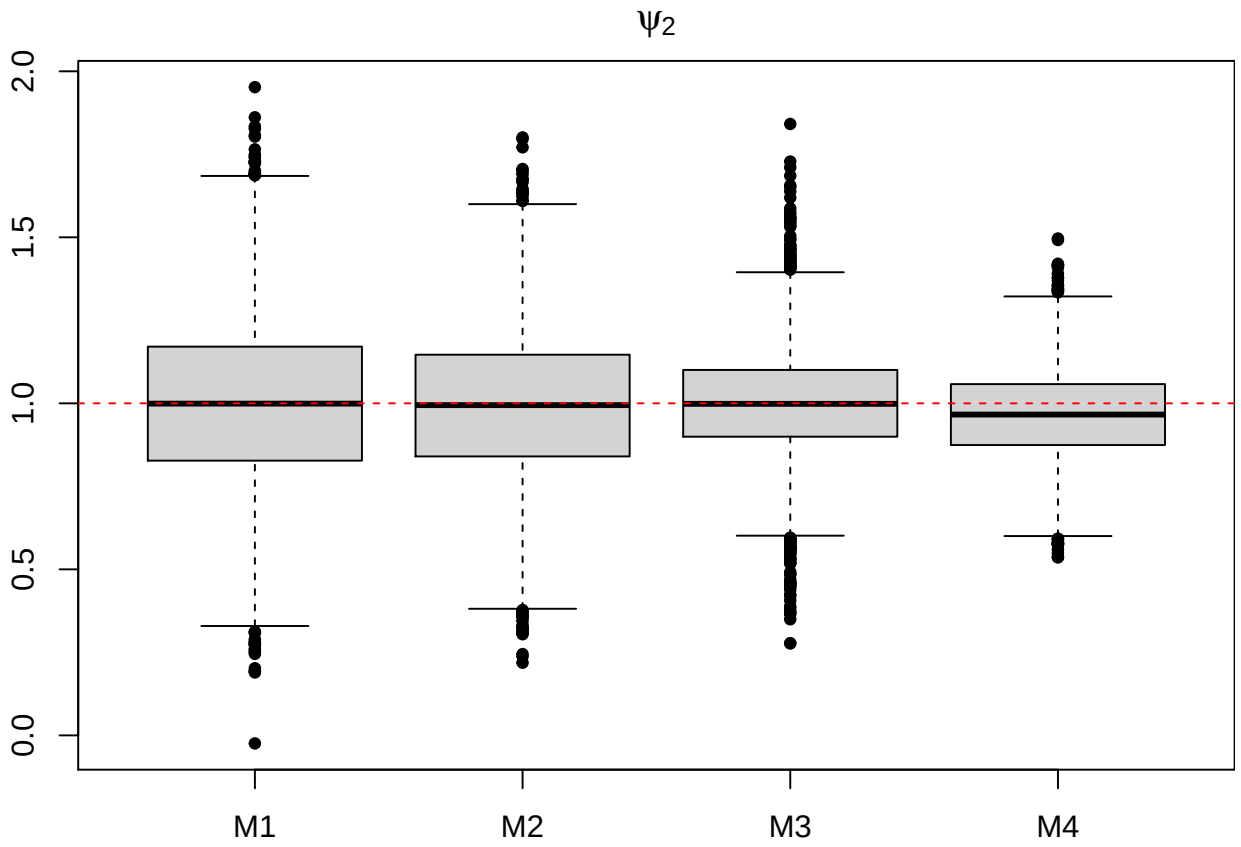
```
set.seed(2987)
n<-2000
nreps<-5000
ests<-array(0,c(nreps,4,4))
for(irep in 1:nreps){
    X<-rmvn(n,mu=Mu,Sigma)
    Xal<-cbind(1,X[,1],X[,2],X[,3],X[,1]*X[,2])
    ps.true<-expit(Xal %*% al)
    Z<-rbinom(n,1,ps.true)
    Xb<-cbind(1,X[,1],X[,2],X[,1]*X[,2])
    mu0<- Xb %*% be
    Xp<-cbind(1,X[,1],X[,2],X[,1]*X[,2])
    mu1<- (Xp %*% psi)
    eta<-mu0 + Z * mu1 ; sig<-0.1
    Y<-rnorm(n,eta,sig)
    X1<-X[,1];X2<-X[,2];X3<-X[,3]
    true.vals<-c(be,psi)[c(1,2,3,5,4,6,7,8)]

    #True PS
    eX<-ps.true
    ests[irep,1,]<-coef(lm(Y~(Z+Z:X1+Z:X2+Z:X1:X2)+(eX+eX:X1+eX:X2+eX:X1:X2)))[c(2,4,5,8)]

    #Fitted PS
    eX<-fitted(glm(Z~X1+X2+X3+X1:X2,family=binomial))
    ests[irep,2,]<-coef(lm(Y~(Z+Z:X1+Z:X2+Z:X1:X2)+(eX+eX:X1+eX:X2+eX:X1:X2)))[c(2,4,5,8)]

    #Fitted PS confounders only
    eX<-fitted(glm(Z~X1+X2+X1:X2,family=binomial))
    ests[irep,3,]<-coef(lm(Y~(Z+Z:X1+Z:X2+Z:X1:X2)+(eX+eX:X1+eX:X2+eX:X1:X2)))[c(2,4,5,8)]

    #Fitted PS mis-specified
    eX<-fitted(glm(Z~X1,family=binomial))
    ests[irep,4,]<-coef(lm(Y~(Z+Z:X1+Z:X2+Z:X1:X2)+(eX+eX:X1+eX:X2+eX:X1:X2)))[c(2,4,5,8)]


}
```

We report boxplots of the estimates over the replications to approximate the sampling distribution, and then numerical summaries of

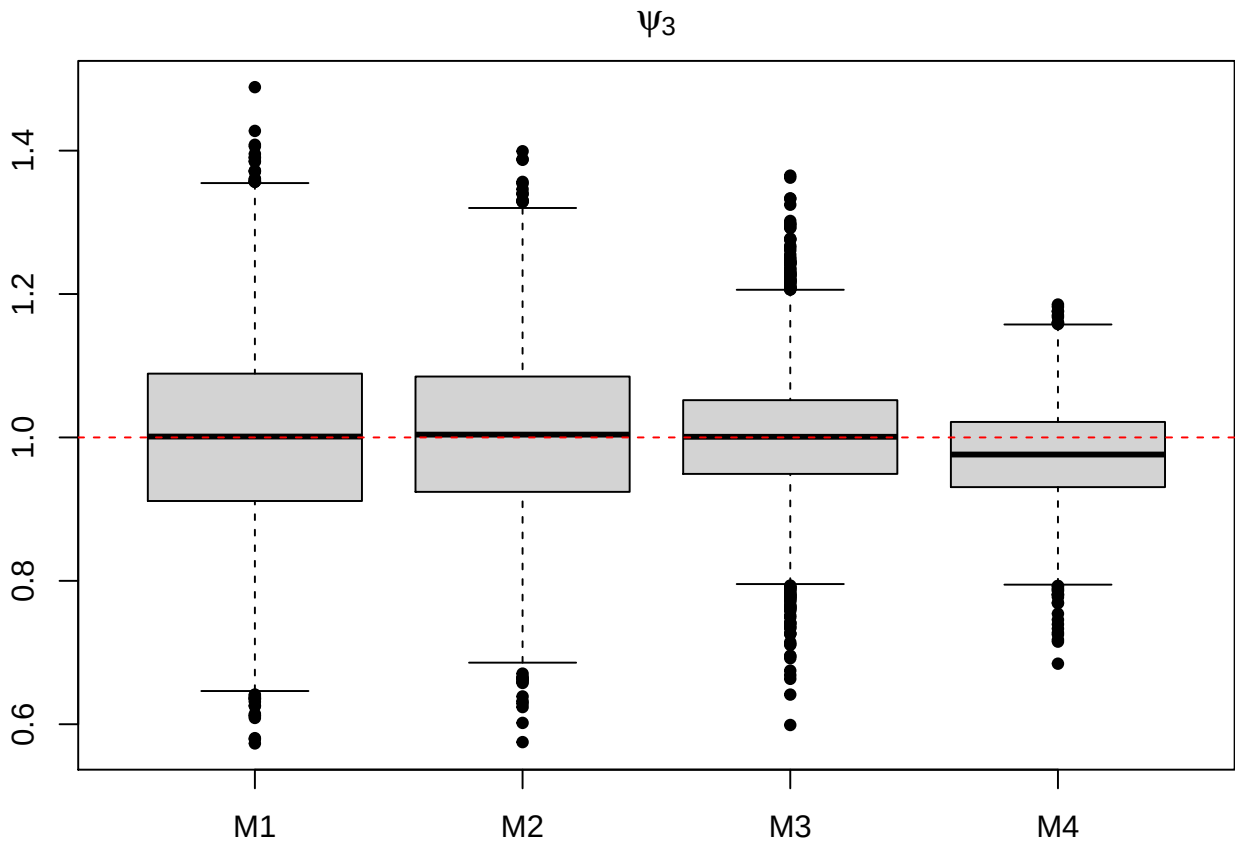
- $\sqrt{n}$ times the *bias* of the estimator
- n times the *variance* of the estimator
- n times the *mean-squared error* of the estimator.



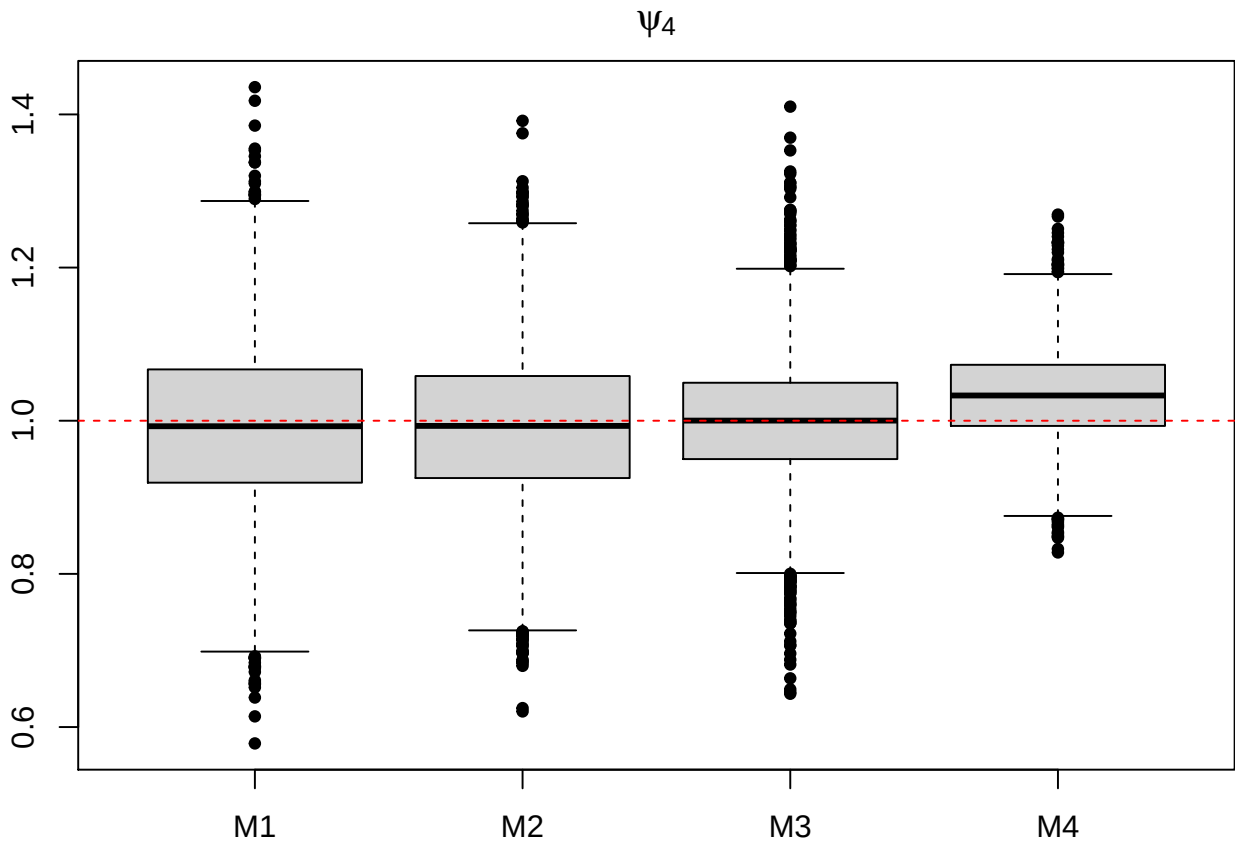

```
## [1] "psi1"
##          Bias      Var.       MSE
## M1 -0.5764555 282.81634 283.14864
## M2 -0.1838973 216.93647 216.97029
## M3  0.1472516  76.71292  76.73460
## M4  2.4383748  56.57341  62.51909
```




```
## [1] "psi2"
##           Bias      Var.       MSE
## M1  0.08075292 127.86493 127.87145
## M2 -0.26522570 102.41474 102.48508
## M3 -0.08616226  53.29948  53.30690
## M4 -1.60606698  37.81356  40.39301
```




```
## [1] "psi3"
##             Bias     Var.      MSE
## M1  0.007970574 33.77638 33.77644
## M2  0.153328058 27.55036 27.57387
## M3  0.052038413 13.75472 13.75743
## M4 -1.059881214  8.97114 10.09449
```




```
## [1] "psi4"
##           Bias     Var.       MSE
## M1 -0.26578327 24.27452 24.345157
## M2 -0.39036261 20.12819 20.280569
## M3 -0.02477913 12.95649 12.957099
## M4  1.48336285  7.18786  9.388225
```

The results are largely as expected: note that the methods that *estimate* the α parameters yield lower variances, that method M3 which utilizes the confounders only has a lower variance than method M2 that uses the true data generating propensity score specification, and that method M4 produces a biased yet low variance estimator.