

MAR 45% missing - Probit

```
# sample MCAR dataset from PUMS
source("../utils/sampleMAR45.R")
n = 3000
missing_col = c(1,3,7,9,10,11)
set.seed(3)

output_list <- sampleMAR45(n)
df <- output_list[['df']]
df_observed <- output_list[['df_observed']]

apply(is.na(df_observed), MARGIN = 2, mean)
```

```
##      VEH      MV      NP      RMSP      ENG      MARHT      SCHL      RACNUM
## 0.4446667 0.0000000 0.4593333 0.0000000 0.0000000 0.0000000 0.4216667 0.0000000
##      AGEP      WKL      PINCP
## 0.4890000 0.4406667 0.4296667
```

Ordinal bayesian nonparametric model

```
source("../probitBayes.R")
N = 40
Mon = 300
B = 300
thin.int = 1
# function(y, N = 40, Mon = 2000, B = 300, thin.int = 5, seed = 0)
output_list <- probitBayesImputation(df_observed, N, Mon, B, thin.int)

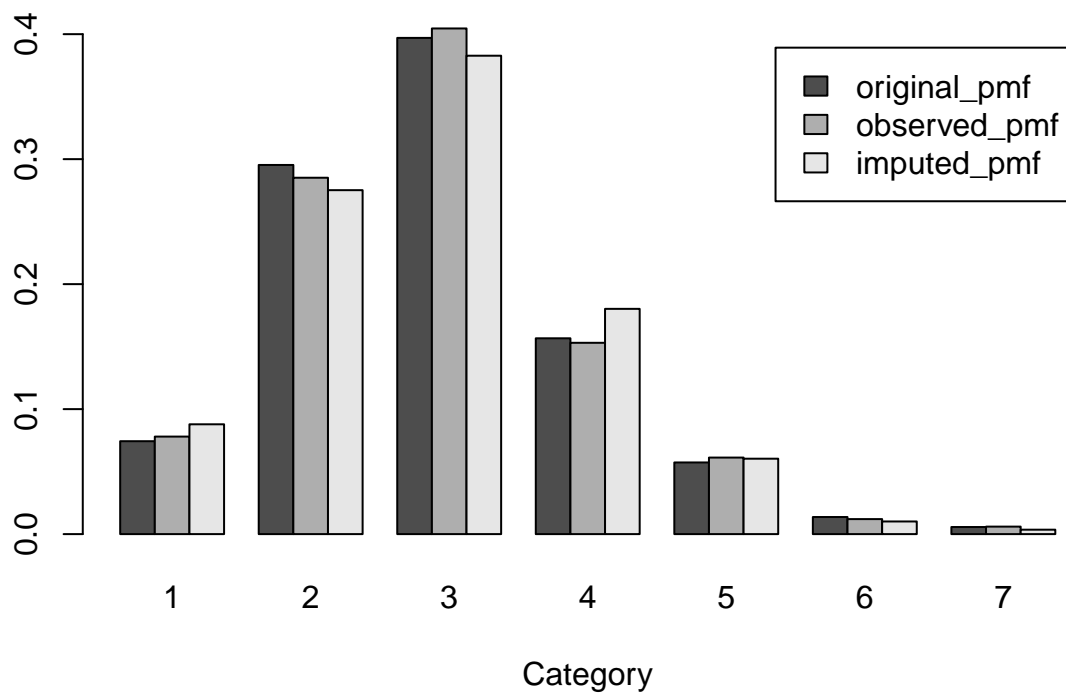
sampled_y <- output_list[['sampled_y']]
sampled_z <- output_list[['sampled_z']]
```

Diagnostics

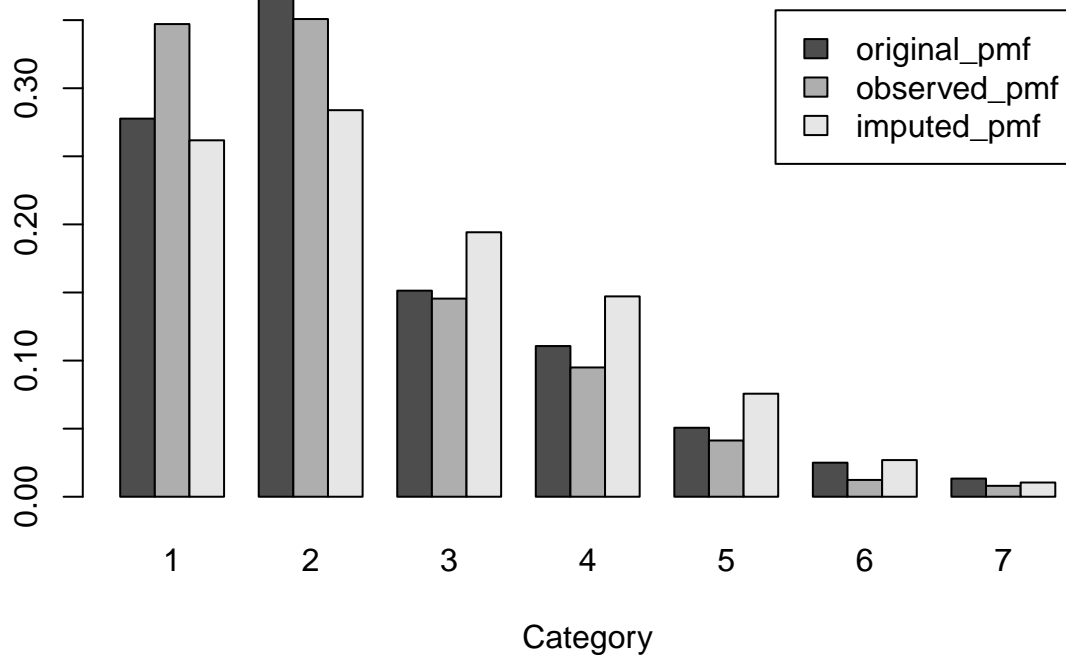
Assess bivariate joint distribution

Assess trivariate joint distribution

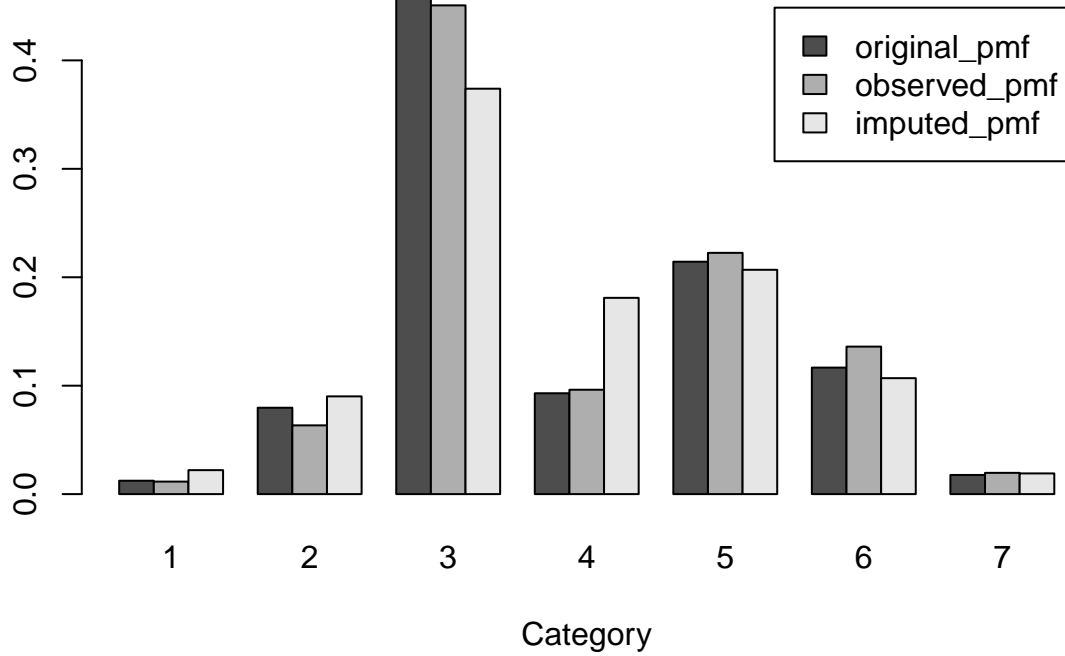
Blocked Gibbs Sampling Assessment: VEH



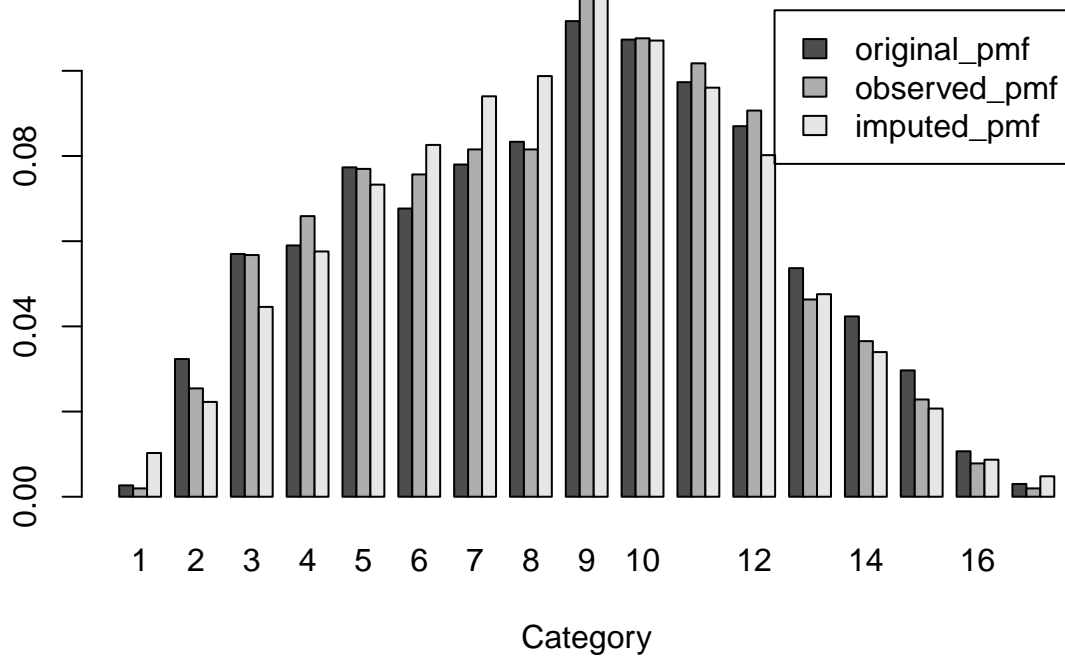
Blocked Gibbs Sampling Assessment: NP



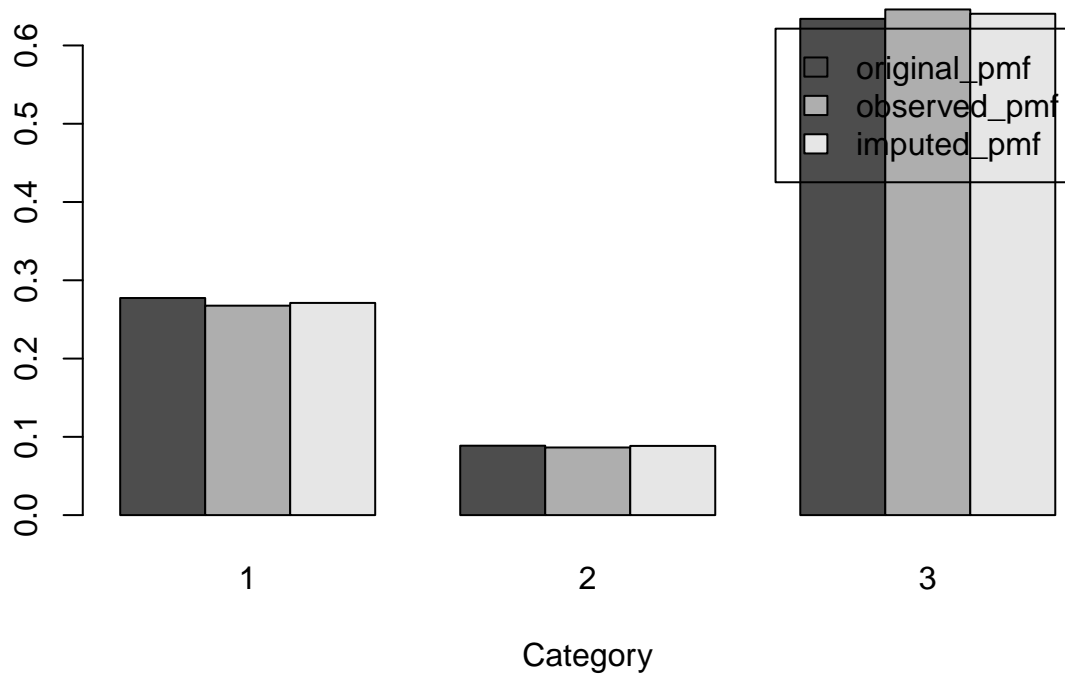
Blocked Gibbs Sampling Assessment: SCHL



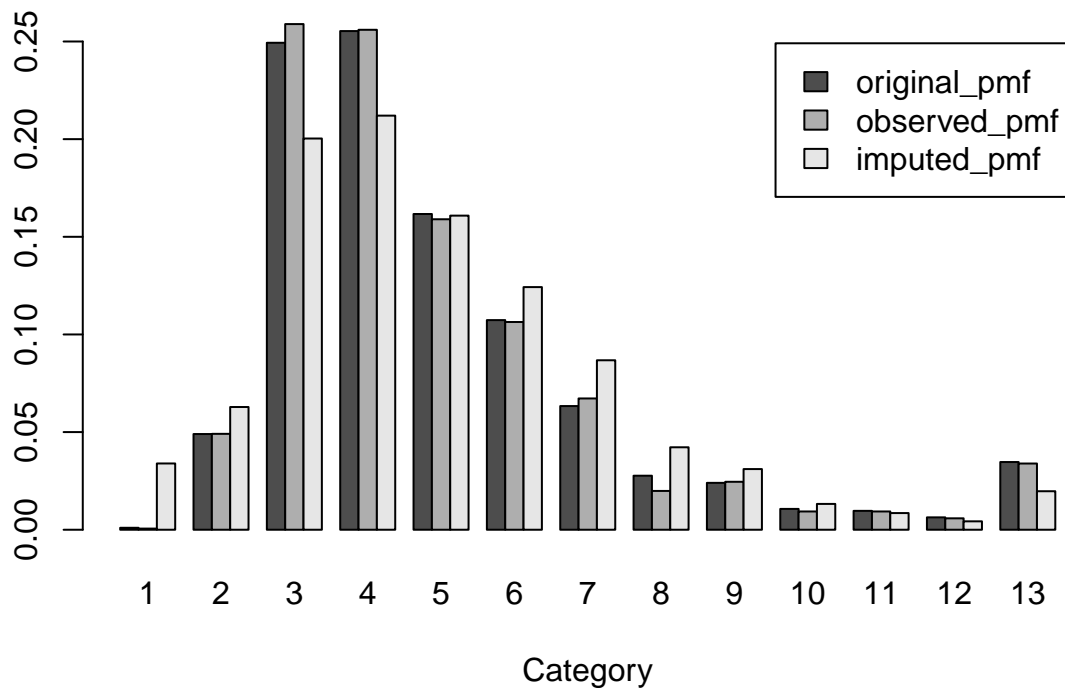
Blocked Gibbs Sampling Assessment: AGEP



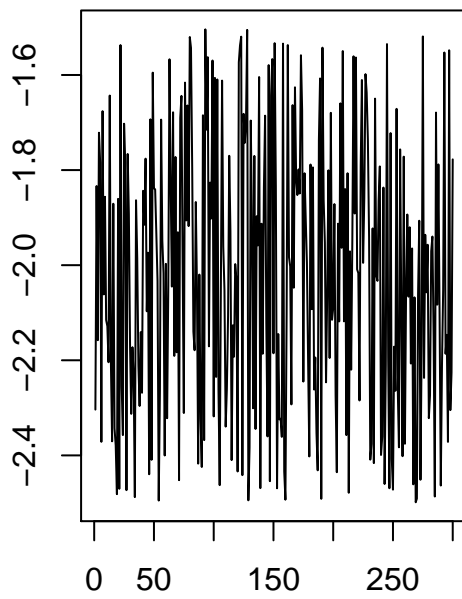
Blocked Gibbs Sampling Assessment: WKL



Blocked Gibbs Sampling Assessment: PINCP

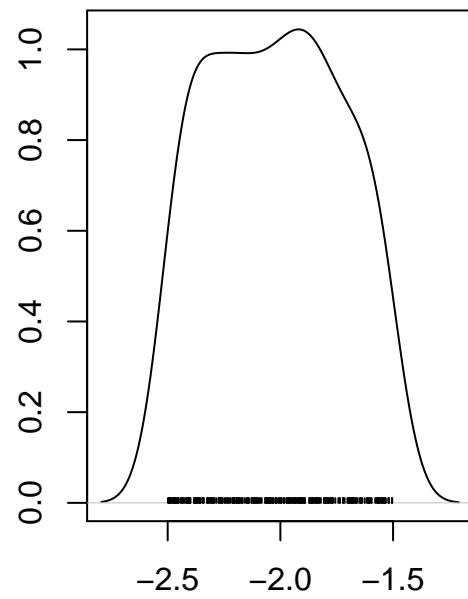


Trace of var1

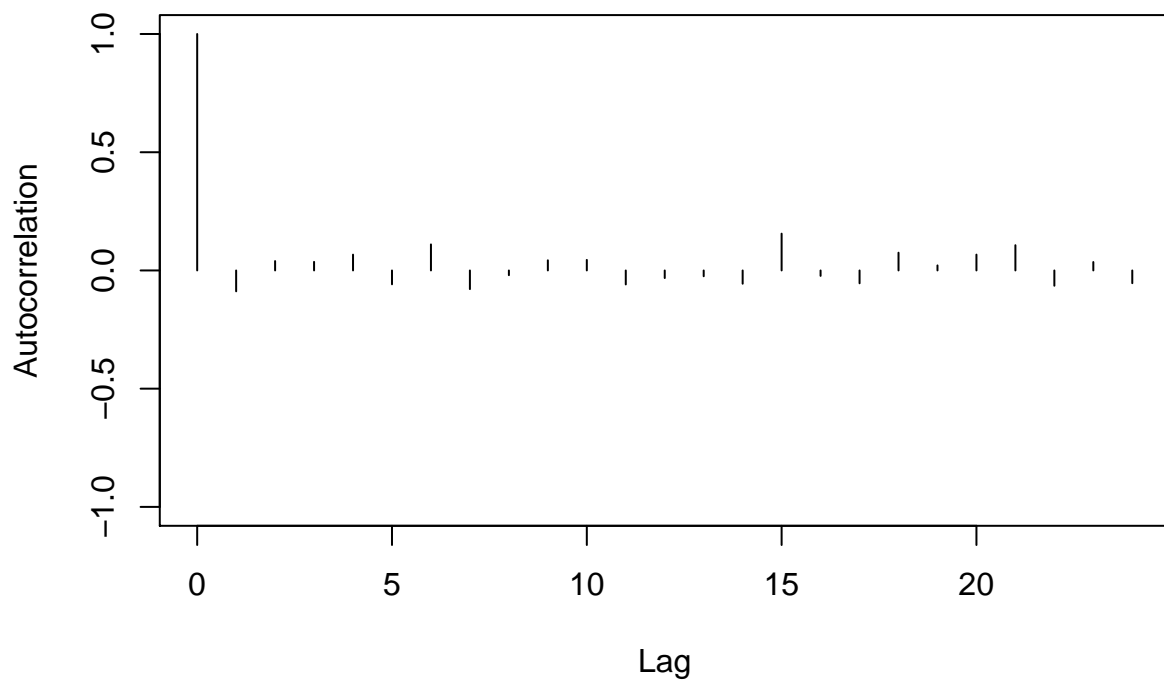


Iterations

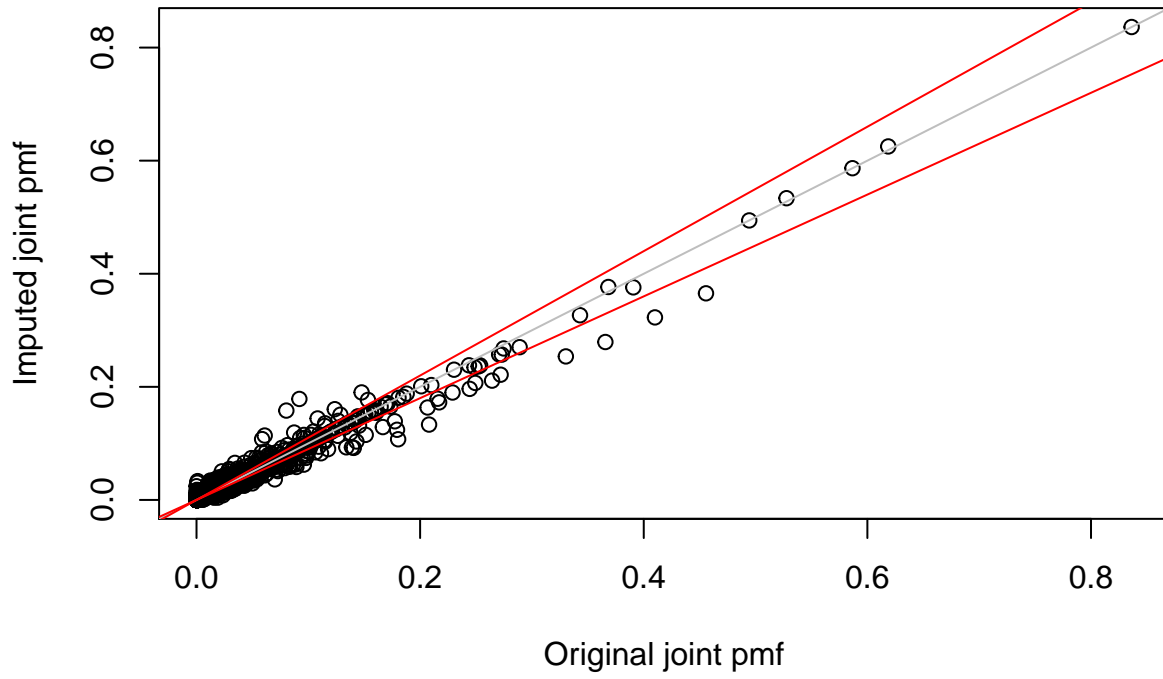
Density of var1



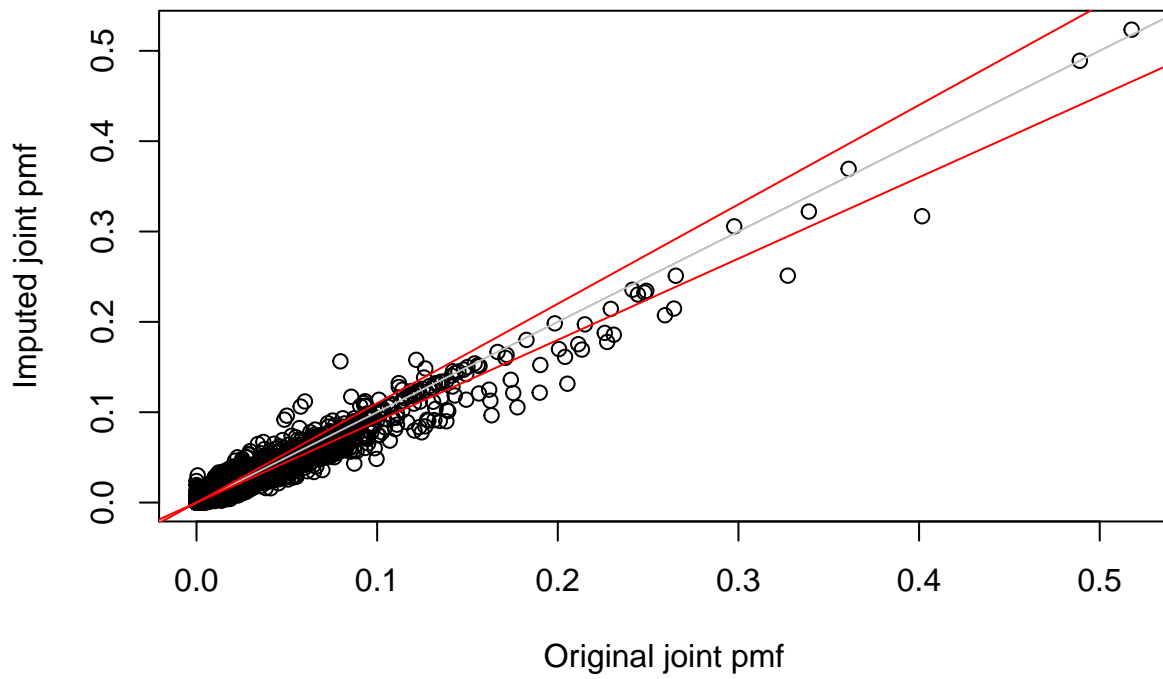
N = 300 Bandwidth = 0.09847



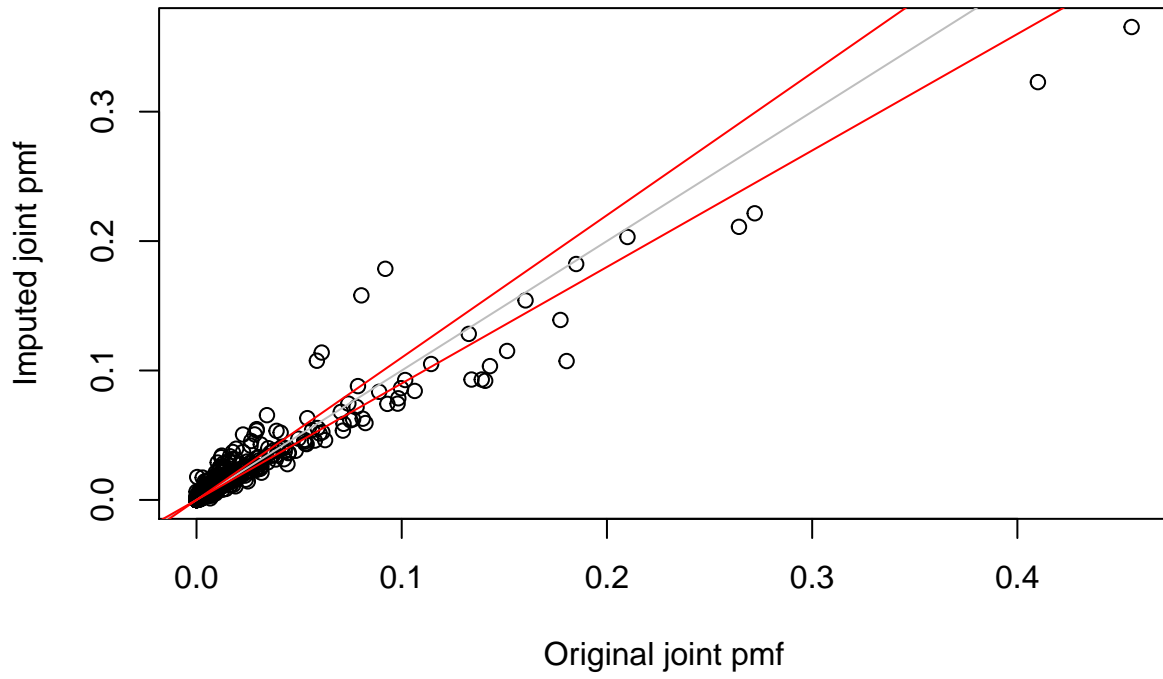
Bivariate pmf



Trivariate pmf



Bivariate pmf SCHL



Bivariate pmf WKL

