

# Testing different imputation methods on PUMS (MCAR)

---

```
# load dataset: df
load('.../Datasets/ordinalPUMS.Rdata')

# take 2000 samples: df
set.seed(0)
n = 3000
sample <- sample(nrow(df), size = n)
df <- df[sample,]

# create MCAR scenario with 30% chance of missing: df_observed
missing_prob = 0.3
df_observed <- df
missing_col = c(1,3,7,9,10,11)
for (col in missing_col) {
  missing_ind <- rbernoulli(n,p = missing_prob)
  df_observed[missing_ind, col] <- NA
}
```

## 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']]
```

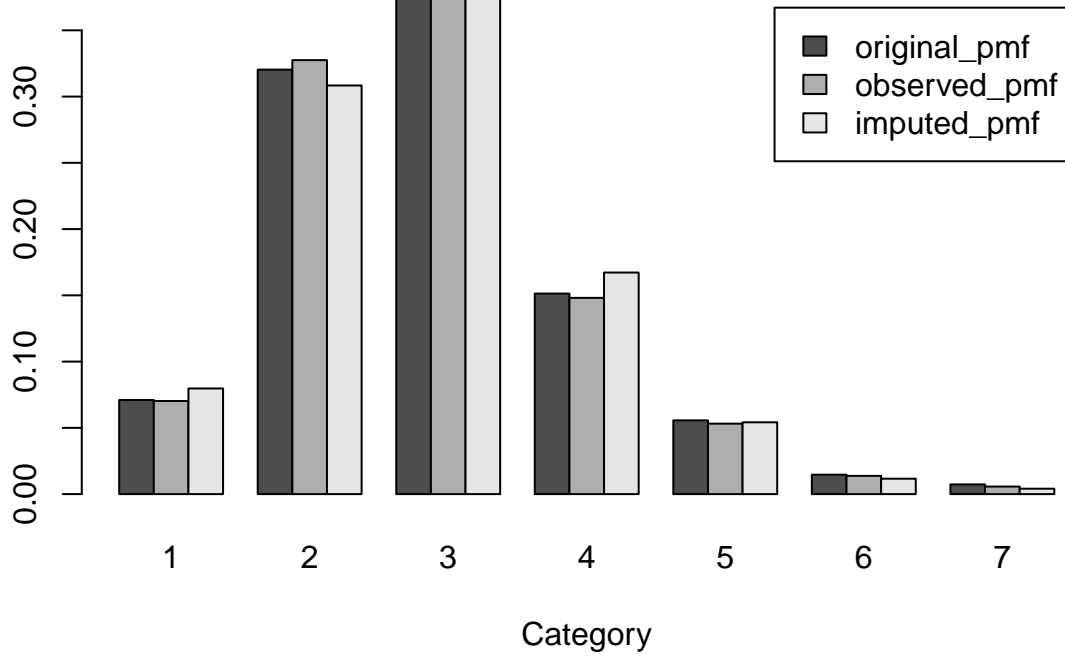
## Diagnostics

Assess bivariate joint distribution

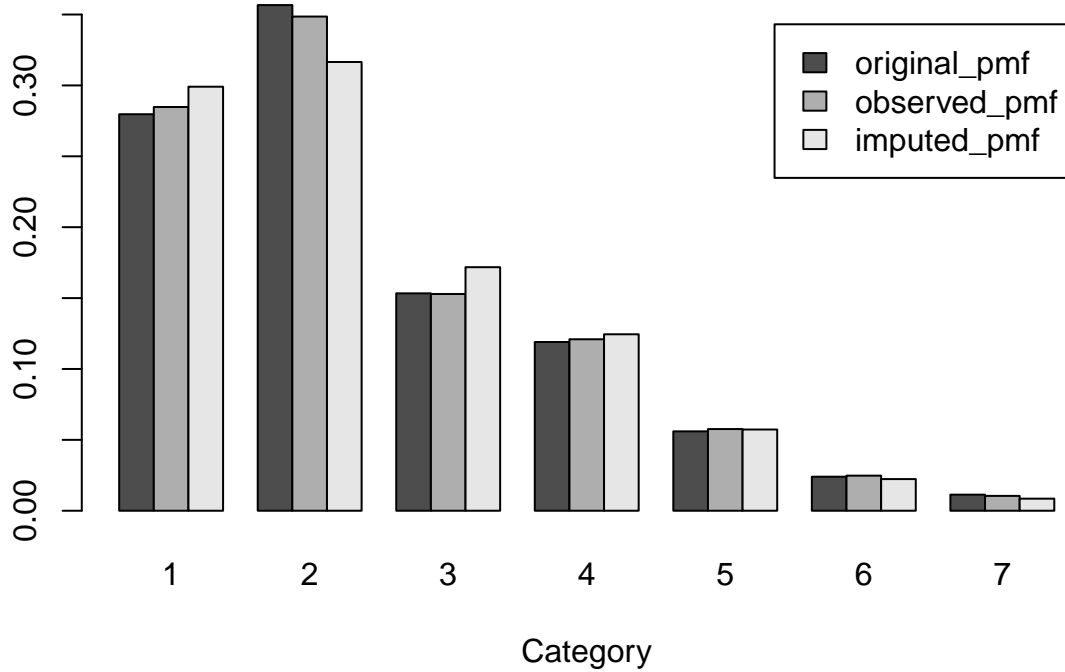
Assess trivariate joint distribution

---

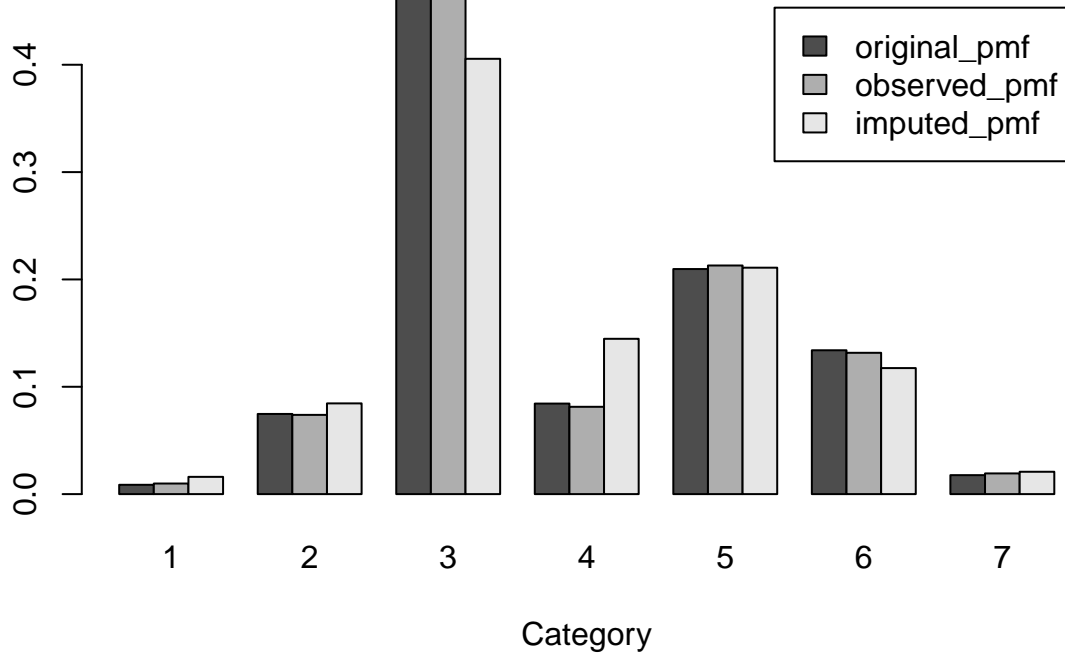
### Blocked Gibbs Sampling Assessment: VEH



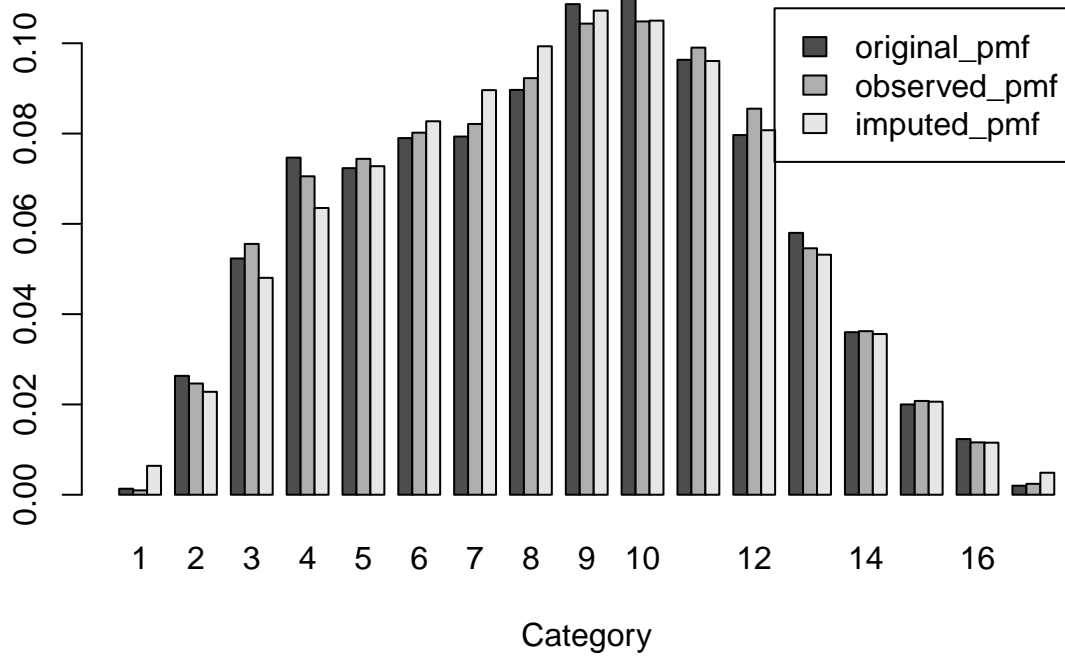
### Blocked Gibbs Sampling Assessment: NP



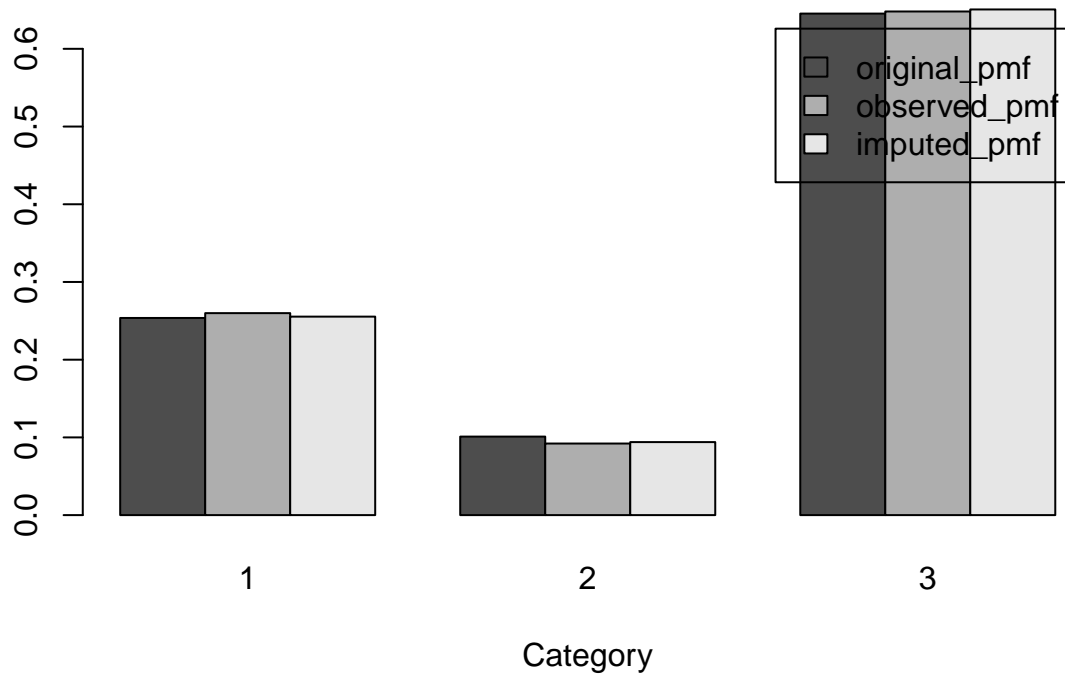
### Blocked Gibbs Sampling Assessment: SCHL



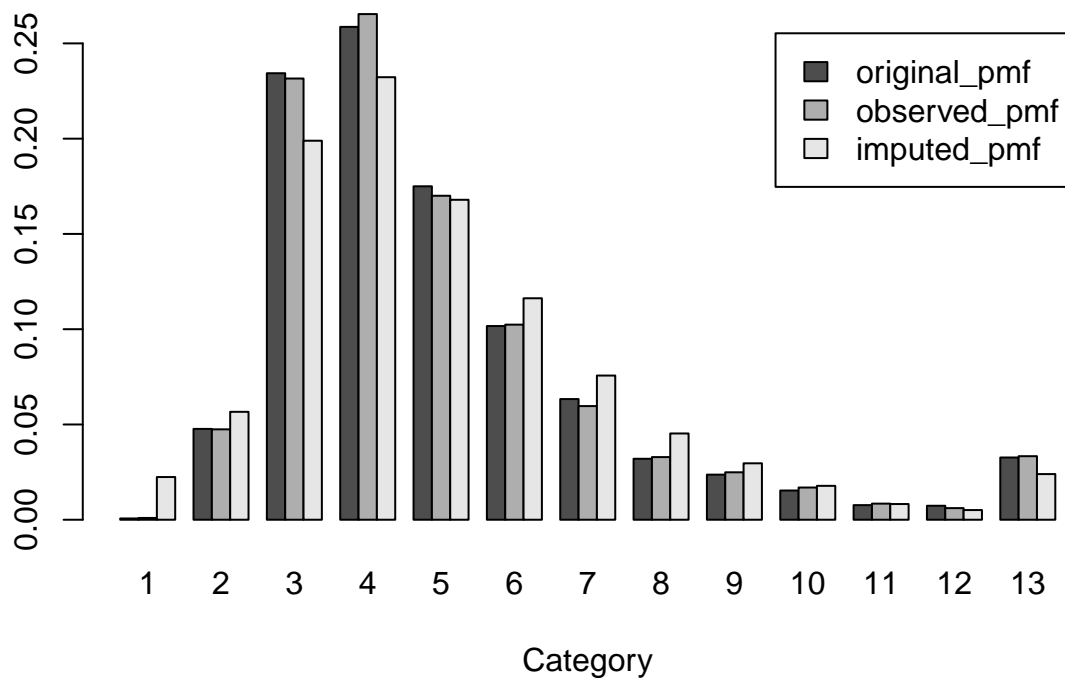
### Blocked Gibbs Sampling Assessment: AGEF



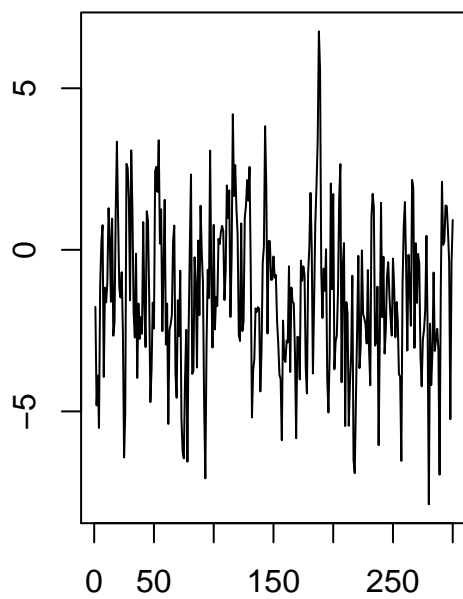
### Blocked Gibbs Sampling Assessment: WKL



### Blocked Gibbs Sampling Assessment: PINCP

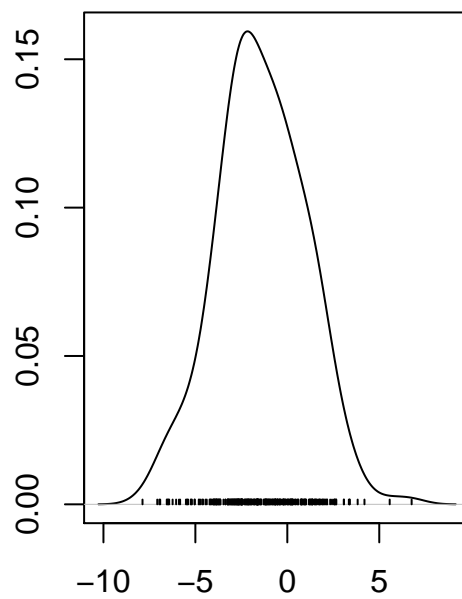


**Trace of var1**

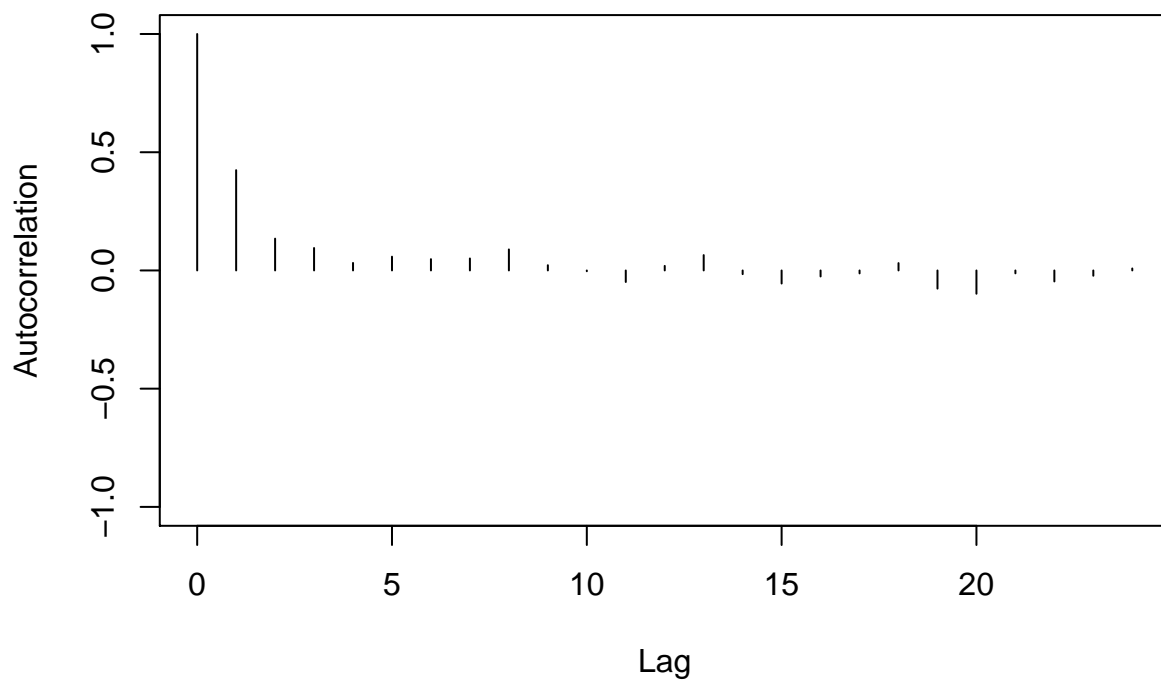


Iterations

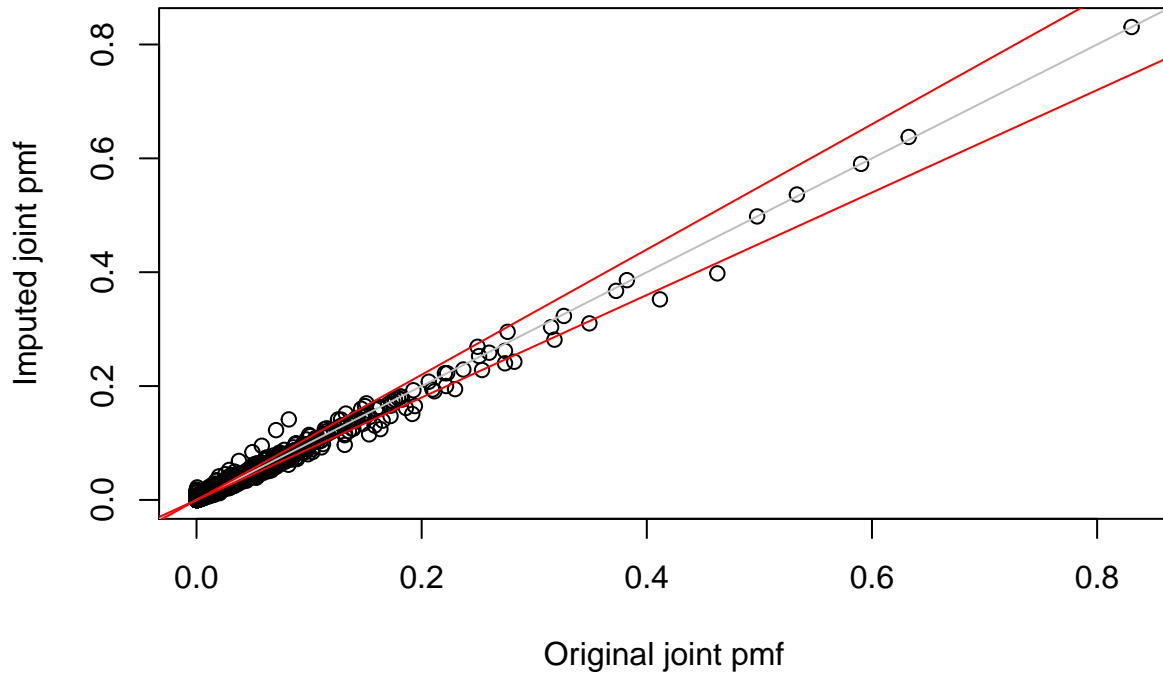
**Density of var1**



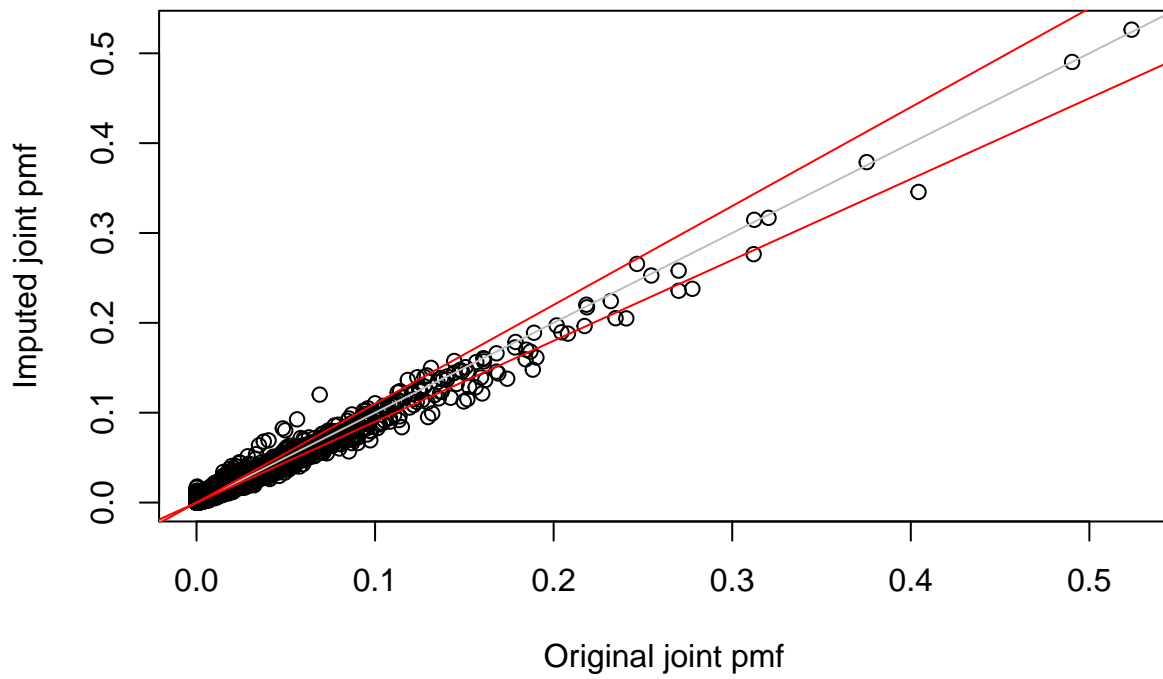
N = 300 Bandwidth = 0.7997



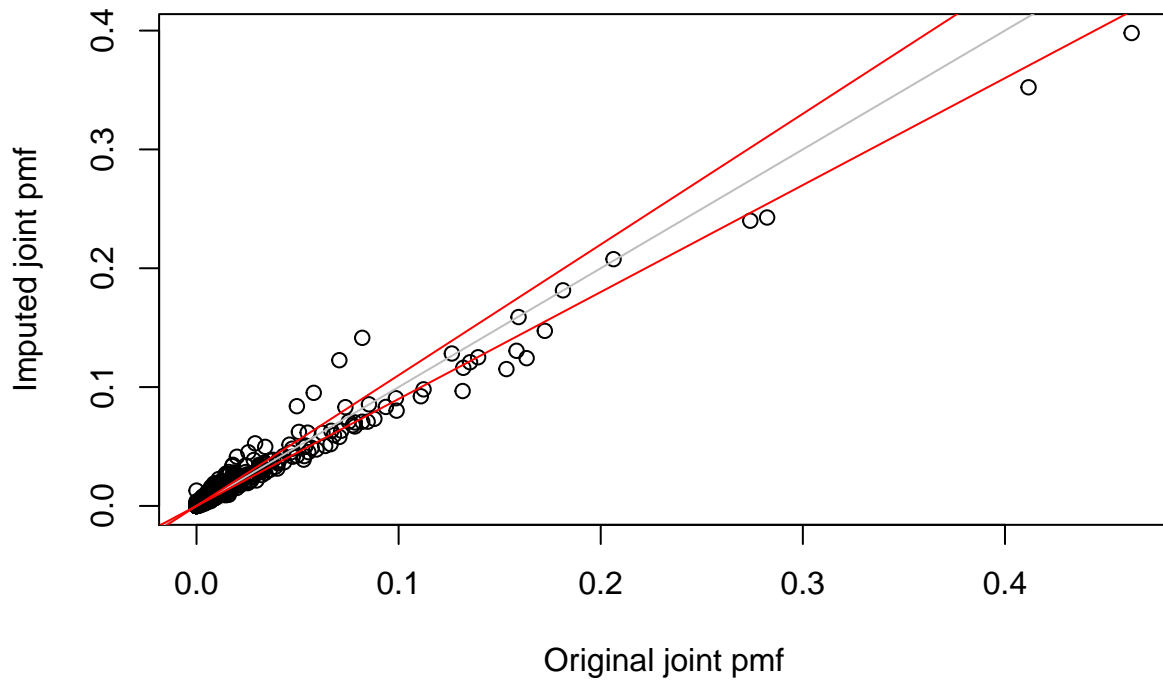
**Bivariate pmf**



**Trivariate pmf**



**Bivariate pmf SCHL**



**Bivariate pmf WKL**

