Paper Part 1: Table 2

Part D: Moment Reconstruction

**> strat.boot4mod**

**function(nboot, y, x1obs, z1, z2, x1ref1, x1ref2)**

**{**

**b1 <- rep(0, nboot)**

**bz1 <- rep(0, nboot)**

**bz2 <- rep(0, nboot)**

**x1ref <- (x1ref1 + x1ref2)/2**

**vare <- 0.5 \* var(x1ref1 - x1ref2, na.method = "omit")**

**y1 <- y[!is.na(x1ref)]**

**x1obs1 <- x1obs[!is.na(x1ref)]**

**z11 <- z1[!is.na(x1ref)]**

**z21 <- z2[!is.na(x1ref)]**

**x1ref11 <- x1ref1[!is.na(x1ref)]**

**x1ref21 <- x1ref2[!is.na(x1ref)]**

**x1ref3 <- x1ref[!is.na(x1ref)]**

**y2 <- y[is.na(x1ref)]**

**x1obs2 <- x1obs[is.na(x1ref)]**

**z12 <- z1[is.na(x1ref)]**

**z22 <- z2[is.na(x1ref)]**

**runexcyz <- glm(x1ref ~ y + z1 + z2, subset = !is.na(x1ref) & !is.na(y))**

**excyz <- runexcyz$coef[1] + runexcyz$coef[2] \* y + runexcyz$coef[3] \* z1 + runexcyz$coef[4] \* z2**

**vxcyz <- runexcyz$deviance/runexcyz$df - vare/2**

**runwcyz <- glm(x1obs ~ y + z1 + z2, subset = !is.na(x1obs) & !is.na(y))**

**ewcyz <- runwcyz$coef[1] + runwcyz$coef[2] \* y + runwcyz$coef[3] \* z1 + runwcyz$coef[4] \* z2**

**vwcyz <- runwcyz$deviance/runwcyz$df**

**xmr <- excyz - sqrt(vxcyz/vwcyz) \* (x1obs - ewcyz)**

**run <- glm(y ~ xmr + z1 + z2, subset = !is.na(xmr) & !is.na(y))**

**print(summary(run))**

**for(i in 1:nboot) {**

**sampk <- sample(seq(length(y1)), replace = T)**

**vare <- 0.5 \* var(x1ref11[sampk] - x1ref21[sampk], na.method = "omit")**

**runexcyz <- glm(x1ref3[sampk] ~ y1[sampk] + z11[sampk] + z21[sampk], subset = !is.na(x1ref3[sampk]) &**

**!**

**is.na(y1[sampk]))**

**excyz <- runexcyz$coef[1] + runexcyz$coef[2] \* y + runexcyz$coef[3] \* z1 + runexcyz$coef[4] \* z2**

**vxcyz <- runexcyz$deviance/runexcyz$df - vare/2**

**sampj <- sample(seq(length(y2)), replace = T)**

**ynew <- c(y1[sampk], y2[sampj])**

**x1obsnew <- c(x1obs1[sampk], x1obs2[sampj])**

**z1new <- c(z11[sampk], z12[sampj])**

**z2new <- c(z21[sampk], z22[sampj])**

**excyz <- runexcyz$coef[1] + runexcyz$coef[2] \* ynew + runexcyz$coef[3] \* z1new + runexcyz$coef[4] \***

**z2new**

**vxcyz <- runexcyz$deviance/runexcyz$df - vare/2**

**runwcyz <- glm(x1obsnew ~ ynew + z1new + z2new, subset = !is.na(x1obsnew) & !is.na(ynew))**

**ewcyz <- runwcyz$coef[1] + runwcyz$coef[2] \* ynew + runwcyz$coef[3] \* z1new + runwcyz$coef[4] \***

**z2new**

**vwcyz <- runwcyz$deviance/runwcyz$df**

**xmr <- excyz - sqrt(vxcyz/vwcyz) \* (x1obsnew - ewcyz)**

**run <- glm(ynew ~ xmr + z1new + z2new, subset = !is.na(xmr) & !is.na(ynew))**

**b1[i] <- run$coef[2]**

**bz1[i] <- run$coef[3]**

**bz2[i] <- run$coef[4]**

**}**

**c(mean(b1), stdev(b1), quantile(b1, probs = seq(0, 1, 0.1)), mean(bz1), stdev(bz1), quantile(bz1, probs =**

**seq(0, 1, 0.1)), mean(bz2), stdev(bz2), quantile(bz2, probs = seq(0, 1, 0.1)))**

**}**

**> strat.boot4mod(5000,bmi,log(pot.den.ffq),age,sex,log(pot.ur1250/dlw1),log(pot.ur2250/dlw1))**

**Call: glm(formula = y ~ xmr + z1 + z2, subset = !is.na(xmr) & !is.na(y))**

**Deviance Residuals:**

**Min 1Q Median 3Q Max**

**-10.30264228 -3.339408748 -0.7214154697 2.302413747 18.61821709**

**Coefficients:**

**Value Std. Error t value**

**(Intercept) 23.69141909692 1.5954048139 14.84978538994**

**xmr -8.13106765948 0.7151429756 -11.36984901813**

**z1 0.10057819271 0.0265463766 3.78877291736**

**z2 0.03567071206 0.4320564625 0.08256030209**

**(Dispersion Parameter for Gaussian family taken to be 21.96114621 )**

**Null Deviance: 13429.12809 on 482 degrees of freedom**

**Residual Deviance: 10519.38903 on 479 degrees of freedom**

**Number of Fisher Scoring Iterations: 1**

**Correlation of Coefficients:**

**(Intercept) xmr z1**

**xmr 0.1867601961**

**z1 -0.9100316581 -0.2344091029**

**z2 -0.4729906803 -0.1196753037 0.0967735301**

**0% 10% 20% 30% 40% 50% 60%**

**-8.471232157 1.768191405 -16.30014252 -10.87512716 -10.00059506 -9.355415991 -8.80152173 -8.307332282 -7.873804493**

**70% 80% 90% 100% 0% 10%**

**-7.426775514 -6.922998218 -6.297616473 -3.633873194 0.1032517139 0.03502119333 -0.01593622145 0.05922780438**

**20% 30% 40% 50% 60% 70% 80% 90%**

**0.07451862535 0.08473202379 0.0934140937 0.1018721821 0.1109914392 0.1213505814 0.1328163385 0.148740967**

**100% 0% 10% 20% 30% 40%**

**0.242880217 0.04347805652 0.5348962422 -2.052094749 -0.6352627342 -0.4059283064 -0.2377852166 -0.1018508814**

**50% 60% 70% 80% 90% 100%**

**0.02951645461 0.1680867227 0.3157877808 0.4841532161 0.7464304043 2.249594653**

>

Part E: Multiple Imputation

> strat.mi

function(nimp, y, x1obs, z1, z2, x1ref, x2ref)

{

b1 <- rep(0, nimp)

bz1 <- rep(0, nimp)

bz2 <- rep(0, nimp)

vb1 <- rep(0, nimp)

vbz1 <- rep(0, nimp)

vbz2 <- rep(0, nimp)

ximp <- rep(0, length(x1obs))

xref <- (x1ref + x2ref)/2

runexcwyz <- glm(xref ~ x1obs + y + z1 + z2, subset = !is.na(xref) & !is.na(y) & !is.na(x1obs))

print(summary(runexcwyz))

a <- runexcwyz$coef

vara <- summary(runexcwyz)$cov

vxrefcwyz <- runexcwyz$deviance/runexcwyz$df

dfvxrefcwyz <- runexcwyz$df

dfvaru <- length(xref[!is.na(xref)])

varx <- var(x1ref, x2ref, na.method = "omit")

varu <- 2 \* (var(xref, na.method = "omit") - varx)

theta1 <- log(vxrefcwyz)

theta2 <- log((vxrefcwyz - 0.5 \* varu)/(0.5 \* varu))

vtheta1 <- 2/dfvxrefcwyz

covtheta12 <- (2 \* vxrefcwyz)/(dfvxrefcwyz \* (vxrefcwyz - 0.5 \* varu))

vtheta2 <- ((2 \* vxrefcwyz^2)/dfvxrefcwyz + varu^2/(2 \* dfvaru))/(vxrefcwyz - 0.5 \* varu)^2 + 2/dfvaru + (2 \*

varu)/(dfvaru \* (vxrefcwyz - 0.5 \* varu))

for(i in 1:nimp) {

a1 <- rmvnorm(1, mean = a, cov = vara)

t1 <- rmvnorm(1, mean = c(theta1, theta2), cov = matrix(c(vtheta1, covtheta12, covtheta12, vtheta2),

nrow = 2))

expt1 <- exp(t1[1, 1])

expt2 <- exp(t1[1, 2])

e1 <- rnorm(length(x1obs), mean = 0, sd = sqrt((expt1 \* expt2)/(1 + expt2)))

ximp1 <- a1[1, 1] + a1[1, 2] \* x1obs + a1[1, 3] \* y + a1[1, 4] \* z1 + a1[1, 5] \* z2

ximp[is.na(xref)] <- ximp1[is.na(xref)] + e1[is.na(xref)]

e2 <- rnorm(length(x1obs), mean = 0, sd = sqrt((expt1 \* expt2)/(1 + expt2)^2))

ximp2 <- ximp1 + ((xref - ximp1) \* expt2)/(1 + expt2)

ximp[!is.na(xref)] <- ximp2[!is.na(xref)] + e2[!is.na(xref)]

runimp <- glm(y ~ ximp + z1 + z2, subset = !is.na(y) & !is.na(ximp))

b1[i] <- runimp$coef[2]

bz1[i] <- runimp$coef[3]

bz2[i] <- runimp$coef[4]

vb1[i] <- summary(runimp)$cov[2, 2]

vbz1[i] <- summary(runimp)$cov[3, 3]

vbz2[i] <- summary(runimp)$cov[4, 4]

}

print(summary(glm(y ~ x1obs + z1 + z2, subset = !is.na(y) & !is.na(x1obs))))

seb1 <- sqrt(mean(vb1) + ((nimp + 1) \* var(b1))/nimp)

sebz1 <- sqrt(mean(vbz1) + ((nimp + 1) \* var(bz1))/nimp)

sebz2 <- sqrt(mean(vbz2) + ((nimp + 1) \* var(bz2))/nimp)

c(mean(b1), stdev(b1), seb1, mean(bz1), stdev(bz1), sebz1, mean(bz2), stdev(bz2), sebz2)

}

> strat.mi(500,bmi,log(pot.den.ffq),age,sex,log(pot.ur1250/dlw1),log(pot.ur2250/dlw1))

Call: glm(formula = xref ~ x1obs + y + z1 + z2, subset = !is.na(xref) & !is.na(y) & !is.na(x1obs))

Deviance Residuals:

Min 1Q Median 3Q Max

-2.032567431 -0.1439849522 0.01935213641 0.1888783473 0.6643523637

Coefficients:

Value Std. Error t value

(Intercept) 0.170761965921 0.184179042655 0.9271519900

x1obs 0.390003412502 0.083210301142 4.6869607146

y -0.023954683956 0.003959868001 -6.0493642598

z1 0.007153328963 0.002458611859 2.9094990886

z2 0.033933640343 0.041705021300 0.8136583866

(Dispersion Parameter for Gaussian family taken to be 0.0898366403 )

Null Deviance: 25.87866777 on 211 degrees of freedom

Residual Deviance: 18.59618453 on 207 degrees of freedom

Number of Fisher Scoring Iterations: 1

Correlation of Coefficients:

(Intercept) x1obs y z1

x1obs -0.1394472211

y -0.5830897097 0.1176066735

z1 -0.6477815473 -0.2019907488 -0.0833038142

z2 -0.4082088045 -0.1224482969 0.0474421889 0.0954737469

Call: glm(formula = y ~ x1obs + z1 + z2, subset = !is.na(y) & !is.na(x1obs))

Deviance Residuals:

Min 1Q Median 3Q Max

-10.23257761 -3.611599501 -0.9421481251 2.613448039 23.6675035

Coefficients:

Value Std. Error t value

(Intercept) 27.27970304926 1.76370401199 15.4672795797

x1obs -1.68861963058 0.93369142715 -1.8085414319

z1 0.03884958636 0.02940883789 1.3210173927

z2 -0.37743182645 0.49133847595 -0.7681707111

(Dispersion Parameter for Gaussian family taken to be 27.69892335 )

Null Deviance: 13429.12809 on 482 degrees of freedom

Residual Deviance: 13267.78429 on 479 degrees of freedom

Number of Fisher Scoring Iterations: 1

Correlation of Coefficients:

(Intercept) x1obs z1

x1obs -0.0628721082

z1 -0.8814587091 -0.1696406259

z2 -0.4397412696 -0.1967040840 0.1021679250

[1] -7.27525623319 2.02361653963 2.03115944182 0.09441454059 0.04721190115 0.04760414310 0.05094720136

[8] 0.72823903777 0.73493738271

>

Paper Part 1: Table 4

Part A: Unadjusted analysis (E(Y|W,Z))

> summary(glm(bmi~log(sod.ffq)+age+sex,subset=!is.na(pot.ffq)))

Call: glm(formula = bmi ~ log(sod.ffq) + age + sex, subset = !is.na(pot.ffq))

Deviance Residuals:

Min 1Q Median 3Q Max

-10.46952355 -3.573006007 -0.9924995566 2.579401266 23.528603

Coefficients:

Value Std. Error t value

(Intercept) 17.34003135602 5.25512536820 3.299641805

log(sod.ffq) 1.12980010748 0.57445755141 1.966725139

age 0.03653680386 0.02916489754 1.252766406

sex -0.23412434022 0.50788356324 -0.460980345

(Dispersion Parameter for Gaussian family taken to be 27.66466679 )

Null Deviance: 13429.12809 on 482 degrees of freedom

Residual Deviance: 13251.37539 on 479 degrees of freedom

Number of Fisher Scoring Iterations: 1

Correlation of Coefficients:

(Intercept) log(sod.ffq) age

log(sod.ffq) -0.9423088243

age -0.4117736648 0.1169802700

sex -0.4466985288 0.3184606113 0.1042849956

>

Part B1: Model of true log sodium biomarker on BMI, sex and age (**E(X|Y,Z) model)**

**summary(glm(logsod.urav[1:250]~bmi[1:250]+age[1:250]+sex[1:250],subset=!is.na(logsod.urav[1:250])))**

**Call: glm(formula = logsod.urav[1:250] ~ bmi[1:250] + age[1:250] + sex[1:250], subset = !is.na(logsod.urav[1:250]))**

**Deviance Residuals:**

**Min 1Q Median 3Q Max**

**-1.375973384 -0.1668219792 0.03063924657 0.1957046701 0.9911208767**

**Coefficients:**

**Value Std. Error t value**

**(Intercept) 8.031575488057 0.184887641597 43.440304710**

**bmi[1:250] 0.030615450959 0.003920546516 7.808975313**

**age[1:250] -0.002716268604 0.002449851371 -1.108748325**

**sex[1:250] -0.286923092016 0.040982468587 -7.001117842**

**(Dispersion Parameter for Gaussian family taken to be 0.0965242234 )**

**Null Deviance: 33.71612541 on 232 degrees of freedom**

**Residual Deviance: 22.10404715 on 229 degrees of freedom**

**Number of Fisher Scoring Iterations: 1**

**Correlation of Coefficients:**

**(Intercept) bmi[1:250] age[1:250]**

**bmi[1:250] -0.5822148456**

**age[1:250] -0.7048511155 -0.0464654455**

**sex[1:250] -0.4322047204 0.0857158881 0.0635580120**

Part B2: Model of FFQ-reported log sodium on BMI, sex and age (**E(W|Y,Z) model)**

**> summary(glm(log(sod.ffq)~bmi+age+sex,subset=!is.na(sod.ffq)))**

**Call: glm(formula = log(sod.ffq) ~ bmi + age + sex, subset = !is.na(sod.ffq))**

**Deviance Residuals:**

**Min 1Q Median 3Q Max**

**-1.399828682 -0.2280483702 -0.002415142298 0.2683346039 1.166635293**

**Coefficients:**

**Value Std. Error t value**

**(Intercept) 8.428224810329 0.170147333998 49.534862594**

**bmi 0.007090180539 0.003605069361 1.966725139**

**age -0.006150502760 0.002297061398 -2.677552618**

**sex -0.277638784589 0.038191077552 -7.269729015**

**(Dispersion Parameter for Gaussian family taken to be 0.1736125539 )**

**Null Deviance: 94.03594482 on 482 degrees of freedom**

**Residual Deviance: 83.16041333 on 479 degrees of freedom**

**Number of Fisher Scoring Iterations: 1**

**Correlation of Coefficients:**

**(Intercept) bmi age**

**bmi -0.5737512316**

**age -0.7152198900 -0.0468111208**

**sex -0.4078112006 0.0521275352 0.0685869360**

>

Part C: Adjusted analysis using moment reconstruction

**> strat.boot4mod(5000,bmi,log(sod.ffq),age,sex,log(sod.ur1250),log(sod.ur2250))**

**Call: glm(formula = y ~ xmr + z1 + z2, subset = !is.na(xmr) & !is.na(y))**

**Deviance Residuals:**

**Min 1Q Median 3Q Max**

**-10.52907387 -2.645165139 -0.5091419915 2.359960528 15.36318644**

**Coefficients:**

**Value Std. Error t value**

**(Intercept) -81.10536858012 6.54870737132 -12.384943162**

**xmr 12.20959412104 0.72205168321 16.909584736**

**z1 0.05184207113 0.02304966026 2.249146866**

**z2 3.15741290113 0.44095758087 7.160355186**

**(Dispersion Parameter for Gaussian family taken to be 17.46344333 )**

**Null Deviance: 13429.12809 on 482 degrees of freedom**

**Residual Deviance: 8364.989355 on 479 degrees of freedom**

**Number of Fisher Scoring Iterations: 1**

**Correlation of Coefficients:**

**(Intercept) xmr z1**

**xmr -0.9769596379**

**z1 -0.2484575717 0.0564839072**

**z2 -0.5715859459 0.4975098425 0.0897673191**

**0% 10% 20% 30% 40% 50% 60%**

**12.6879816 2.69694389 4.221543515 9.610154443 10.50679942 11.20047148 11.78181647 12.40325345 13.03418297**

**70% 80% 90% 100% 0% 10%**

**13.77740161 14.67517206 15.99388374 30.20415868 0.05283910366 0.03492025103 -0.0873370447 0.009737129187**

**20% 30% 40% 50% 60% 70% 80% 90%**

**0.02374801919 0.03447378323 0.04329711971 0.05207266265 0.06112765738 0.07073691002 0.08146649131 0.09803052382**

**100% 0% 10% 20% 30% 40% 50%**

**0.236694796 3.293380777 0.9989113451 0.6399223615 2.10034697 2.472157387 2.757694549 2.975849215 3.214770714**

**60% 70% 80% 90% 100%**

**3.452445847 3.72057038 4.046083941 4.583794369 10.20712722**

Table 5: Part A: Unadjusted analysis

See Table 4: Part A above.

Table 5: Part B: Model of true log sodium biomarker on log FFQ sodium, BMI, sex and age

and Part C: Adjusted analysis using multiple imputation (see end of output)

> strat.mi(500,bmi,log(sod.ffq),age,sex,log(sod.ur1250),log(sod.ur2250))

Call: glm(formula = xref ~ x1obs + y + z1 + z2, subset = !is.na(xref) & !is.na(y) & !is.na(x1obs))

Deviance Residuals:

Min 1Q Median 3Q Max

-1.441851036 -0.1713426424 0.03002821444 0.1870092736 0.9966239455

Coefficients:

Value Std. Error t value

(Intercept) 7.535647717250 0.445176106427 16.9273409072

x1obs 0.059421858516 0.048533979150 1.2243351886

y 0.029741220855 0.003980844709 7.4710829047

z1 -0.002246292062 0.002477109746 -0.9068197584

z2 -0.269458398813 0.043351939516 -6.2156019274

(Dispersion Parameter for Gaussian family taken to be 0.096314352 )

Null Deviance: 33.71612541 on 232 degrees of freedom

Residual Deviance: 21.95967226 on 228 degrees of freedom

Number of Fisher Scoring Iterations: 1

Correlation of Coefficients:

(Intercept) x1obs y z1

x1obs -0.9098844647

y -0.0744151855 -0.1793701574

z1 -0.4298826797 0.1549639229 -0.0729555629

z2 -0.4687113351 0.3290425765 0.0206096293 0.1102834988

Call: glm(formula = y ~ x1obs + z1 + z2, subset = !is.na(y) & !is.na(x1obs))

Deviance Residuals:

Min 1Q Median 3Q Max

-10.46952355 -3.573006007 -0.9924995566 2.579401266 23.528603

Coefficients:

Value Std. Error t value

(Intercept) 17.34003135602 5.25512536820 3.299641805

x1obs 1.12980010748 0.57445755141 1.966725139

z1 0.03653680386 0.02916489754 1.252766406

z2 -0.23412434022 0.50788356324 -0.460980345

(Dispersion Parameter for Gaussian family taken to be 27.66466679 )

Null Deviance: 13429.12809 on 482 degrees of freedom

Residual Deviance: 13251.37539 on 479 degrees of freedom

Number of Fisher Scoring Iterations: 1

Correlation of Coefficients:

(Intercept) x1obs z1

x1obs -0.9423088243

z1 -0.4117736648 0.1169802700

z2 -0.4466985288 0.3184606113 0.1042849956

[1] 11.02188598210 2.64209993874 2.65030548581 0.05067150039 0.06983981236 0.07013488495 2.75296236433

[8] 1.44323902710 1.44859168401

>