

1 Appendix 1: STATA code used for data cleaning

```
2      #STATA code used for data cleaning
3
4      codebook v024
5      gen region = v024
6      codebook v025
7      gen residence = v025
8      codebook v130
9      replace v130=1 if v130==99
10     gen religion=.
11     replace religion = 1 if v130==7
12     replace religion = 1 if v130==96
13     replace religion = 2 if v130==6
14     replace religion = 3 if v130==5
15     replace religion = 3 if v130==4
16     replace religion = 3 if v130==3
17     replace religion = 3 if v130==2
18     replace religion = 3 if v130==1
19     codebook v131
20     replace v131=1 if v131==99
21     gen ethnicity=.
22     replace ethnicity=1 if v131==13
23     replace ethnicity=1 if v131==96
24     replace ethnicity=1 if v131==2
25     replace ethnicity=1 if v131==4
26     replace ethnicity=1 if v131==7
27     replace ethnicity=1 if v131==8
28     replace ethnicity=1 if v131==10
29     replace ethnicity=1 if v131==11
30     replace ethnicity=2 if v131==3
31     replace ethnicity=2 if v131==5
32     replace ethnicity=2 if v131==6
33     replace ethnicity=3 if v131==1
34     replace ethnicity=3 if v131==9
35     replace ethnicity=3 if v131==12
36     codebook v190
37     gen wealth=.
38     replace wealth=1 if v190==1
39     replace wealth=1 if v190==2
40     replace wealth=2 if v190==3
41     replace wealth=3 if v190==4
42     replace wealth=3 if v190==5
43     gen current_age=v012
44     gen marital_status=.
45     replace marital_status=1 if v501==0
46     replace marital_status=2 if v501==1
47     replace marital_status=2 if v501==2
48     replace marital_status=3 if v501==3
49     replace marital_status=3 if v501==4
50     replace marital_status=3 if v501==5
51     gen schooling=v133
52     gen fertility=v201
```

```

53 codebook v364
54 gen modern_contraceptive=.
55 replace modern_contraceptive=1 if v364==4
56 replace modern_contraceptive=1 if v364==3
57 replace modern_contraceptive=1 if v364==2
58 replace modern_contraceptive=2 if v364==1
59 codebook v717
60 gen occupation=.
61 replace occupation=1 if v717==0
62 replace occupation=2 if v717==3
63 replace occupation=2 if v717==4
64 replace occupation=2 if v717==6
65 replace occupation=2 if v717==9
66 replace occupation=2 if v717==8
67 replace occupation=3 if v717==1
68 replace occupation=3 if v717==2
69 replace occupation=3 if v717==7
70 replace occupation=3 if v717==5
71 gen age_at_1st_sex = v531
72 replace age_at_1st_sex =rpoisson(16.59018) if age_at_1st_sex >=50
73 replace age_at_1st_sex =rpoisson(16.59018) if age_at_1st_sex <=5
74 hist age_at_1st_sex
75 gen schooling_grpd=.
76 replace schooling_grpd=1 if schooling==0
77 replace schooling_grpd=2 if schooling==1
78 replace schooling_grpd=2 if schooling==2
79 replace schooling_grpd=2 if schooling==3
80 replace schooling_grpd=2 if schooling==4
81 replace schooling_grpd=2 if schooling==5
82 replace schooling_grpd=2 if schooling==6
83 replace schooling_grpd=2 if schooling==7
84 replace schooling_grpd=2 if schooling==8
85 replace schooling_grpd=3 if schooling > 8
86 gen fertility_grpd=.
87 replace fertility_grpd=1 if fertility==0
88 replace fertility_grpd=2 if fertility==1
89 replace fertility_grpd=2 if fertility==2
90 replace fertility_grpd=2 if fertility==3
91 replace fertility_grpd=2 if fertility==4
92 replace fertility_grpd=3 if fertility>=5
93 gen region2=.
94 replace region2=1 if region==2
95 replace region2=0 if region2==.
96 gen region3=.
97 replace region3=1 if region==3
98 replace region3=0 if region3==.
99 gen residence2=.
100 replace residence2=1 if residence==2
101 replace residence2=0 if residence2==.
102 gen ethnicity2=.
103 replace ethnicity2=1 if ethnicity==2
104 replace ethnicity2=0 if ethnicity2==.

```

```

105     gen ethnicity3=.
106     replace ethnicity3=1 if ethnicity==3
107     replace ethnicity3=0 if ethnicity3==.
108     gen wealth2=.
109     replace wealth2=1 if wealth==2
110     replace wealth2=0 if wealth2==.
111     gen wealth3=.
112     replace wealth3=1 if wealth==3
113     replace wealth3=0 if wealth3==.
114     gen maritalstatus2=.
115     replace maritalstatus2=1 if marital_status==2
116     replace maritalstatus2=0 if maritalstatus2==.
117     gen maritalstatus3=.
118     replace maritalstatus3=1 if marital_status==3
119     replace maritalstatus3=0 if maritalstatus3==.
120     gen contraceptive2=.
121     replace contraceptive2=1 if modern_contraceptive==2
122     replace contraceptive2=0 if contraceptive2==.
123     gen occupation2=.
124     replace occupation2=1 if occupation==2
125     replace occupation2=0 if occupation2==.
126     gen occupation3=.
127     replace occupation3=1 if occupation==3
128     replace occupation3=0 if occupation3==.
129     gen religion2=.
130     replace religion2=1 if religion==2
131     replace religion2=0 if religion2==.
132     gen religion3=.
133     replace religion3=1 if religion==3
134     replace religion3=0 if religion3==.
135

```

136 2 Appendix 2: R code used for bivariate Poisson 137 model fitting, outlier and correlation analyses

```

138     #R code for bivariate Poisson model fitting, outlier and correlation
139     ↪ analyses
140
141
142     rm(list=ls())
143     library(car)
144     library(foreign)
145     library(VGAMdata)
146     library(ggplot2)
147     library(reshape2)
148     library(nnet)
149     library(ggrepel)
150     library(dplyr)
151     library(data.table)
152     library(readstata13)
153     library(ggpubr)
154     require(foreign)

```

```

155
156 mydata = read.dta("C:/Users/User/Documents/rdata2015.dta",convert.
157     ↪ factors=F)
158
159 model1 <- vglm(cbind(schooling, fertility) ~ religion2 +religion3 +
160     ↪ region2 + region3+residence2+ethnicity2+ethnicity3+wealth2+
161     ↪ wealth3+current_age+maritalstatus2+maritalstatus3+
162     ↪ contraceptive2+occupation2+occupation3+age_at_1st_sex,
163     ↪ poissonff, data = mydata)
164
165 summary(model1)
166
167 AIC = -2*-50279.06 + 2*34
168
169 cor.test(mydata$schooling,mydata$fertility,method="spearman")
170
171 mydata$theta_sch = exp(coef(model1)[1]+mydata$religion2*(coef(model1)
172     ↪ ) [3]) +mydata$religion3*(coef(model1)[5]) +mydata$region2*(
173     ↪ coef(model1)[7]) + mydata$region3*(coef(model1)[9])+mydata$
174     ↪ residence2*(coef(model1)[11])+mydata$ethnicity2*(coef(model1)
175     ↪ [13])+mydata$ethnicity3*(coef(model1)[15])+mydata$wealth2*(
176     ↪ coef(model1)[17])+mydata$wealth3*(coef(model1)[19])+mydata$
177     ↪ current_age*(coef(model1)[21])+mydata$maritalstatus2*(coef(
178     ↪ model1)[23])+mydata$maritalstatus3*(coef(model1)[25])+mydata$
179     ↪ contraceptive2*(coef(model1)[27])+mydata$occupation2*(coef(
180     ↪ model1)[29])+mydata$occupation3*(coef(model1)[31])+mydata$age
181     ↪ _at_1st_sex*(coef(model1)[33]))
182
183 mydata$theta_tfr = exp(coef(model1)[2]+mydata$religion2*(coef(model1)
184     ↪ ) [4]) +mydata$religion3*(coef(model1)[6]) +mydata$region2*(
185     ↪ coef(model1)[8]) + mydata$region3*(coef(model1)[10])+mydata$
186     ↪ residence2*(coef(model1)[12])+mydata$ethnicity2*(coef(model1)
187     ↪ [14])+mydata$ethnicity3*(coef(model1)[16])+mydata$wealth2*(
188     ↪ coef(model1)[18])+mydata$wealth3*(coef(model1)[20])+mydata$
189     ↪ current_age*(coef(model1)[22])+mydata$maritalstatus2*(coef(
190     ↪ model1)[24])+mydata$maritalstatus3*(coef(model1)[26])+mydata$
191     ↪ contraceptive2*(coef(model1)[28])+mydata$occupation2*(coef(
192     ↪ model1)[30])+mydata$occupation3*(coef(model1)[32])+mydata$age
193     ↪ _at_1st_sex*(coef(model1)[34]))
194
195 cor.test(mydata$theta_sch,mydata$theta_tfr,method="spearman")
196
197 corr1 <- ggplot(mydata,aes(x=mydata$theta_sch,y=mydata$theta_tfr))+
198     ↪ geom_point(pch=1)+stat_cor(method="spearman")+labs(x="
199     ↪ estimated_years_of_schooling",y="estimated_number_of_children
200     ↪ born")+theme_test()
201
202 ggsave('test.tiff', corr1, device = "tiff", units="in", width=6,
203     ↪ height=4, pointsize=12,bg="white", dpi = 1200)
204
205 mydata$r1 = mydata$schooling - mydata$theta_sch
206

```

```

207 mydata$r2 = mydata$fertility - mydata$theta_tfr
208
209 mydata$sgn1=ifelse(mydata$r1 > 0,1,-1)
210
211 mydata$sgn2=ifelse(mydata$r2 > 0,1,-1)
212
213 mydata$d1=mydata$sgn1*sqrt(2*(mydata$schooling*log((mydata$schooling
214   ↪ +0.000001)/mydata$theta_sch)-mydata$r1))
215
216 mydata$d2=mydata$sgn2*sqrt(2*(mydata$fertility*log((mydata$fertility
217   ↪ +0.000001)/mydata$theta_tfr)-mydata$r2))
218
219 mydata$Di = (mydata$d1 + mydata$d2)/2
220
221 mydata$id = 1:nrow(mydata)
222
223 outlierII04 <- ggplot(mydata,aes(x=mydata$id,y=mydata$Di))+geom_
224   ↪ point(pch=1)+labs(x="woman_id",y="outlier_estimates_to_
225   ↪ bivariate_Poisson_model")+geom_text(aes(label = mydata$id),
226   ↪ size = 2.8, vjust = 0.5)+ theme_test()+geom_hline(yintercept
227   ↪ = c(-2.58,2.58),color=c('red','red'))
228
229 ggsave('test.tiff', outlierII04, device = "tiff", units="in", width=
230   ↪ 6, height=4.5, pointsize=12,bg="white", dpi = 1200)
231
232 write.table(mydata,file="newMWIR2004.csv",sep="," ,row.names = F)
233

```
