653.11.23

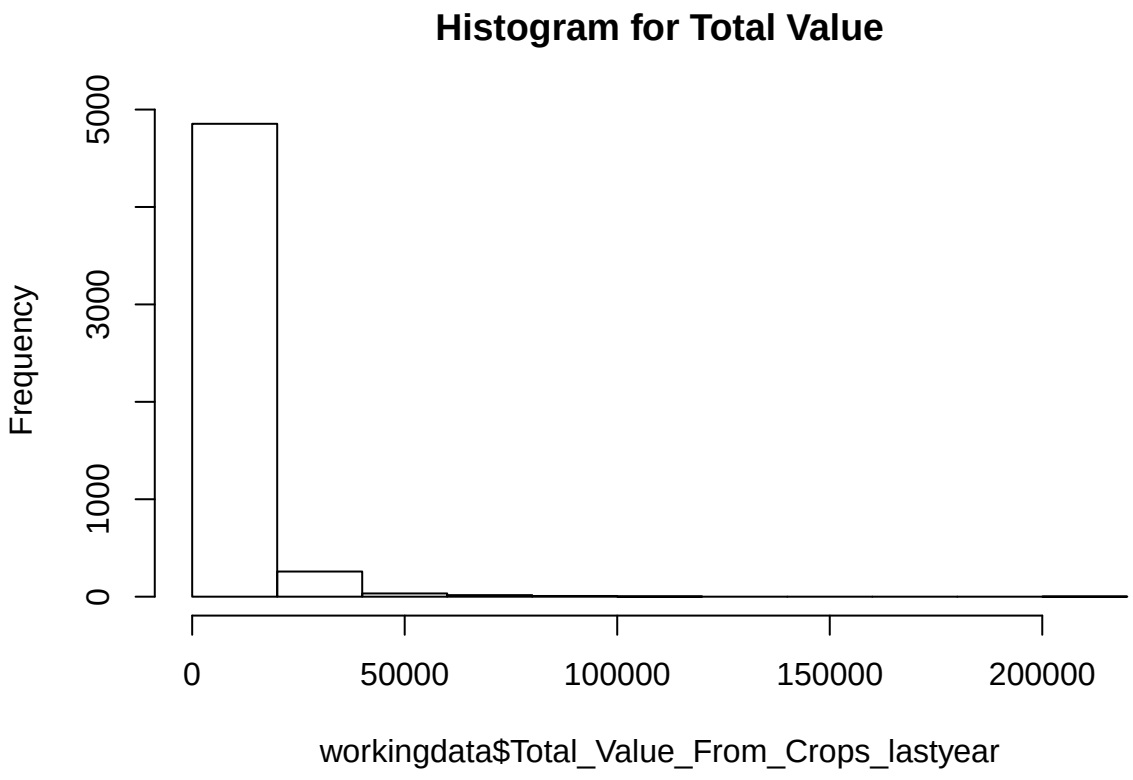
Yibo Wang, Qianhui Huang, Tongbo Xu, Sunyi Chi, Yi Zhao

11/23/2019

Pre-processing

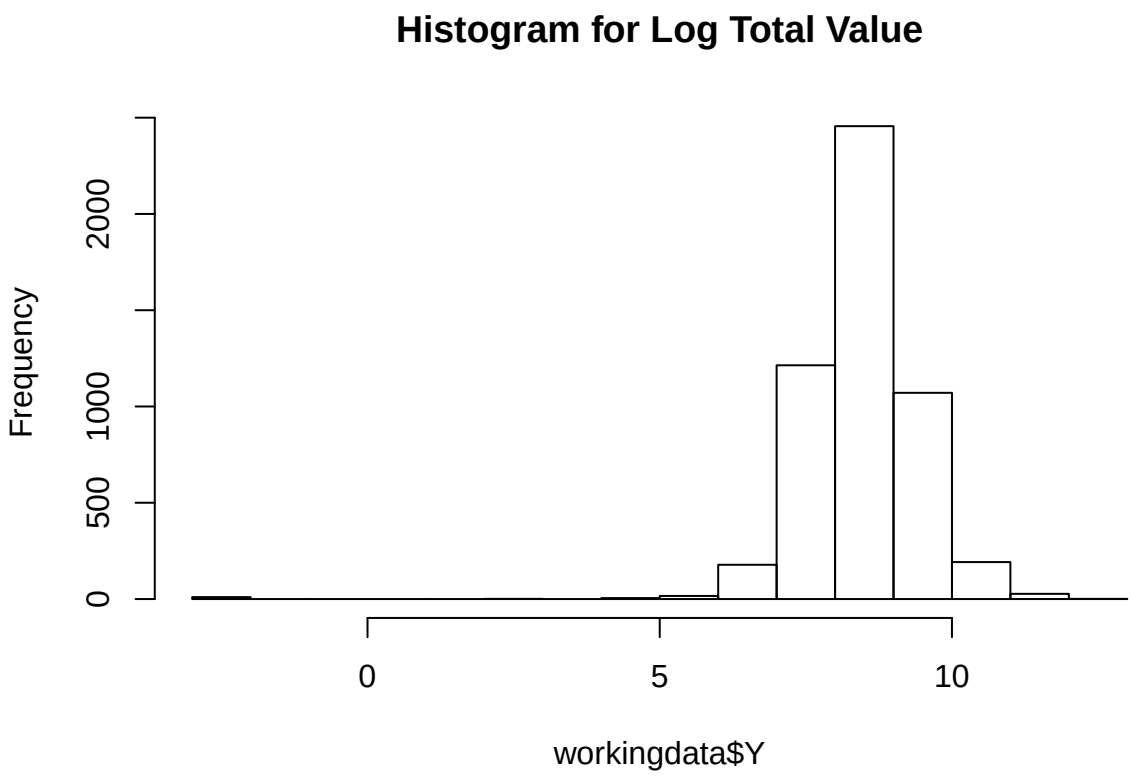
```
load("~/Desktop/653_project/Master_Agriculture_201804/workingdata.Rdata")
workingdata[,c("Year")]<-workingdata[,c("Year")]-2004

workingdata$time<-1
for (i in 1:nrow(workingdata)) {
if (workingdata$Year[i]==2){
  workingdata$time[i]<-2
}
if (workingdata$Year[i]==5){
  workingdata$time[i]<-3
}
if (workingdata$Year[i]==7){
  workingdata$time[i]<-4
}
if (workingdata$Year[i]==11){
  workingdata$time[i]<-5
}}
hist(workingdata$Total_Value_From_Crops_lastyear,main="Histogram for Total Value")
```



```r
workingdata$ratio<-workingdata$Yuan_Spent_Raising_Crops_lastyear/
  (workingdata$Number_of_mu_Land_Cultivated_lastyear+0.01)

workingdata$logmu<-log(workingdata$Number_of_mu_Land_Cultivated_lastyear+0.1)
workingdata$logratio<-log(workingdata$ratio+0.1)
workingdata$Y<-log(workingdata$Total_Value_From_Crops_lastyear+0.1)
hist(workingdata$Y,main="Histogram for Log Total Value")
```




```r
names(workingdata)
```

```
##  [1] "Number_of_mu_Land_Cultivated_lastyear"
##  [2] "Total_Income_From_Crops_lastyear"
##  [3] "Value_of_Crops_Consumed_lastyear"
##  [4] "Yuan_Spent_Raising_Crops_lastyear"
##  [5] "Year"
##  [6] "HHid"
##  [7] "Province"
##  [8] "UR"
##  [9] "Total_Value_From_Crops_lastyear"
## [10] "CI"
## [11] "time"
## [12] "ratio"
## [13] "logmu"
## [14] "logratio"
## [15] "Y"
```

LM fitting

```
lm1<-lm(Y~Year+logmu+logratio+CI,data=workingdata)
summary(lm1)
```

```
##
## Call:
## lm(formula = Y ~ Year + logmu + logratio + CI, data = workingdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.8695  -0.3103   0.0341   0.3306   4.2423
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.578790   0.063496  87.860  < 2e-16 ***
## Year        0.018486   0.002876   6.429  1.4e-10 ***
## logmu       0.766192   0.011172  68.584  < 2e-16 ***
## logratio    0.312838   0.010517  29.747  < 2e-16 ***
## CIInland    0.054520   0.020447   2.666  0.00769 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7141 on 5166 degrees of freedom
## Multiple R-squared:  0.4956, Adjusted R-squared:  0.4952
## F-statistic:  1269 on 4 and 5166 DF,  p-value: < 2.2e-16
```

```
AIC(lm1)
```

```
## [1] 11198.72
```

1. Mean model Selection

1. Check random effect necessity

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
full<-lmer(Y~Year+logmu+logratio+UR+CI+(Year|HHid),data=workingdata,REML=FALSE)
reduce<-lmer(Y~Year+logmu+logratio+UR+CI+(1|HHid),data=workingdata,REML=FALSE)
summary(full)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Y ~ Year + logmu + logratio + UR + CI + (Year | HHid)
##    Data: workingdata
##
```

```
##      AIC      BIC   logLik deviance df.resid
##  10853.6  10919.1  -5416.8  10833.6     5161
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -13.7817  -0.4171   0.0425   0.4532   5.7649
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  HHid     (Intercept) 0.024957 0.15798
##           Year        0.005014 0.07081  -0.78
##  Residual             0.393726 0.62748
## Number of obs: 5171, groups:  HHid, 1360
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  5.629958   0.062014  90.785
## Year         0.018559   0.003329   5.575
## logmu        0.747159   0.011361  65.767
## logratio     0.309437   0.010199  30.341
## URUrban     -0.015405   0.037096  -0.415
## CIInland     0.047270   0.021478   2.201
##
## Correlation of Fixed Effects:
##          (Intr) Year   logmu  lograt URUrbn
## Year     -0.026
## logmu    -0.461  0.015
## logratio -0.924 -0.181  0.253
## URUrban  -0.009  0.002  0.087 -0.027
## CIInland -0.233 -0.019 -0.168  0.121 -0.227
## convergence code: 0
## Model failed to converge with max|grad| = 0.0463368 (tol = 0.002, component 1)
```

```
summary(reduce)
```

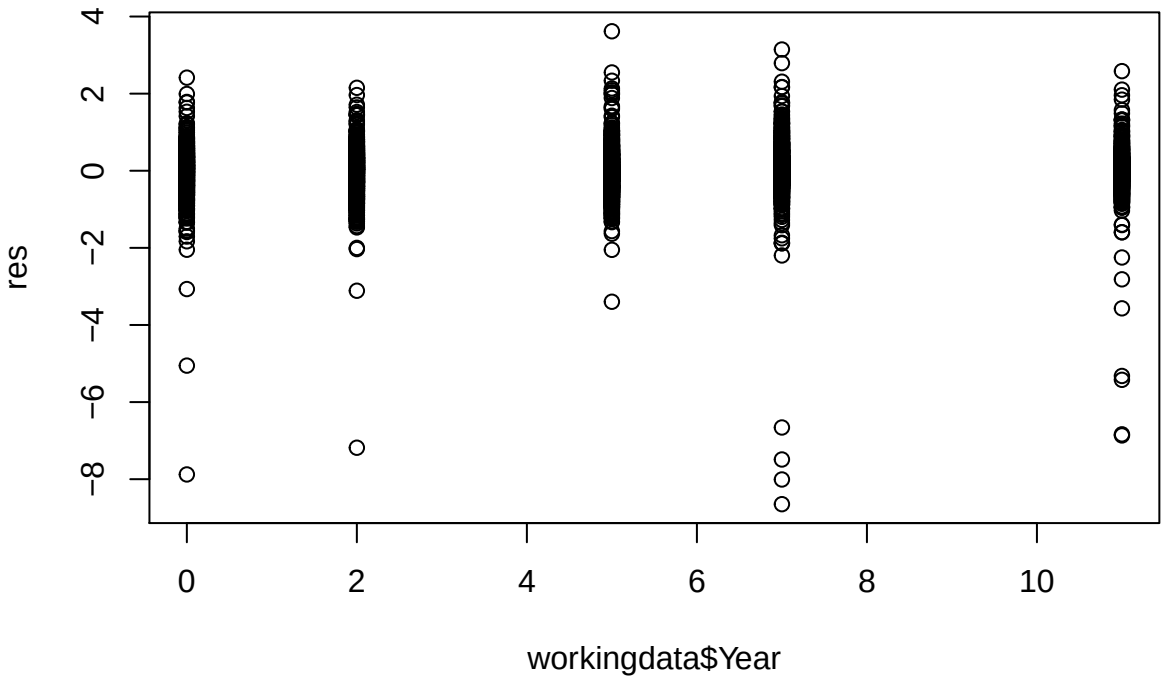
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Y ~ Year + logmu + logratio + UR + CI + (1 | HHid)
##    Data: workingdata
##
##      AIC      BIC   logLik deviance df.resid
##  11163.1  11215.5  -5573.6  11147.1     5163
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -16.2649  -0.4089   0.0438   0.4437   5.6647
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  HHid     (Intercept) 0.03781  0.1944
##  Residual             0.47148  0.6866
## Number of obs: 5171, groups:  HHid, 1360
##
## Fixed effects:
##              Estimate Std. Error t value
```

```
## (Intercept)  5.596619   0.064546  86.707
## Year         0.018658   0.002803   6.657
## logmu        0.764868   0.011974  63.879
## logratio     0.309744   0.010566  29.314
## URUrban     -0.012238   0.039643  -0.309
## CIInland     0.054336   0.023041   2.358
##
## Correlation of Fixed Effects:
##          (Intr) Year   logmu  lograt URUrbn
## Year     -0.018
## logmu    -0.458  0.033
## logratio -0.917 -0.213  0.247
## URUrban  -0.015 -0.004  0.103 -0.027
## CIInland -0.223 -0.006 -0.176  0.104 -0.218
```

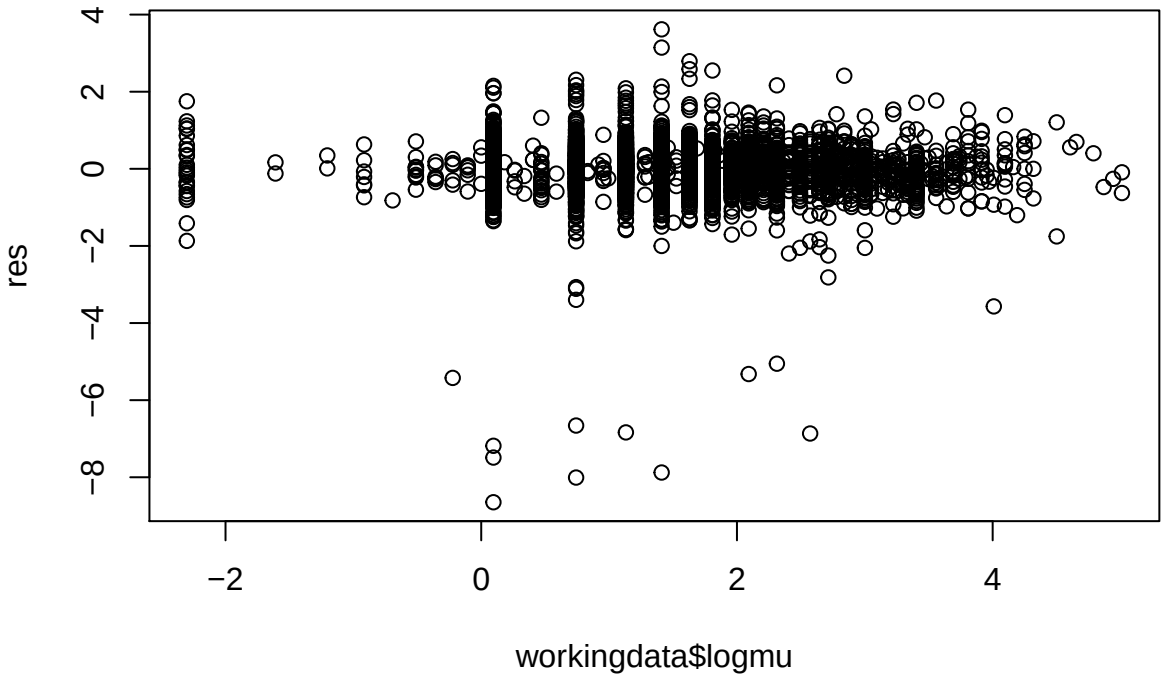
2. Residual, Interaction checking

1. Residual Plot

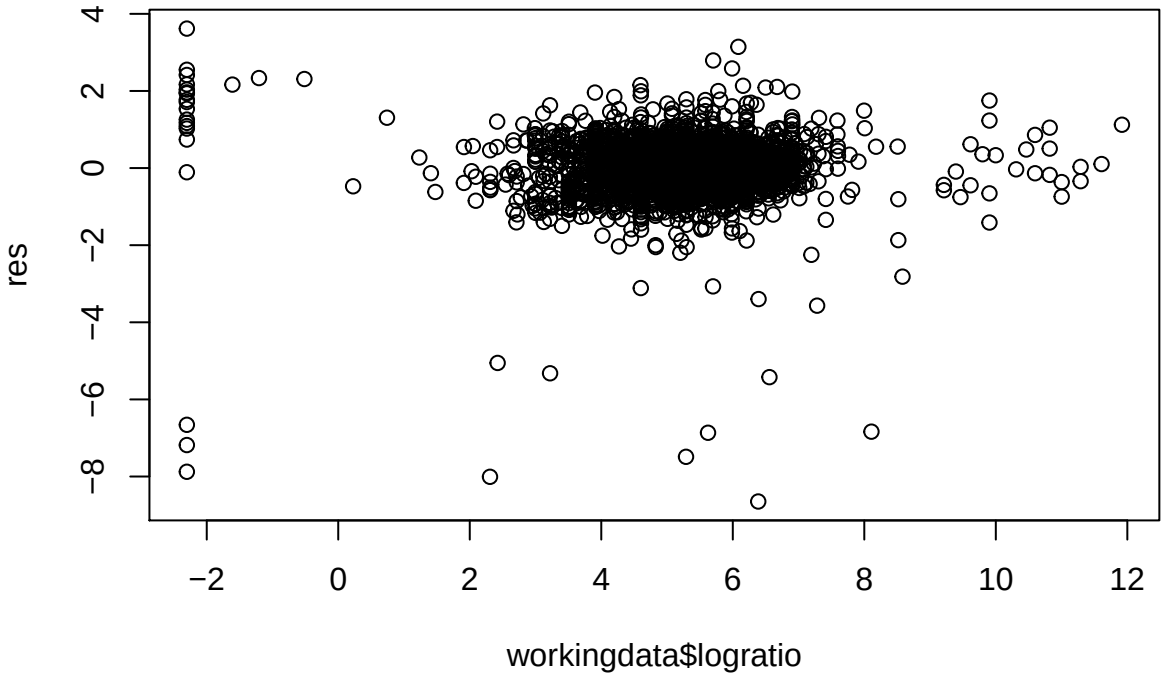
```
res<-(workingdata[,15]-predict(full,workingdata))
plot(workingdata$Year,res)
```




```
plot(workingdata$logmu,res)
```




```
plot(workingdata$logratio,res)
```



2. Interaction

```
int1<-lmer(Y~Year+logmu+logratio+UR+CI+UR*logmu+(Year|HHid),data=workingdata,REML=FALSE)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
```

```
## control$checkConv, : Model failed to converge with max|grad| = 0.0854849
## (tol = 0.002, component 1)
```

```
summary(int1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Y ~ Year + logmu + logratio + UR + CI + UR * logmu + (Year |
##     HHid)
##    Data: workingdata
##
##      AIC      BIC   logLik deviance df.resid
##  10854.5  10926.6  -5416.3  10832.5     5160
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -13.8193  -0.4203   0.0426   0.4542   5.7664
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  HHid     (Intercept) 0.025028 0.15820
##           Year        0.005018 0.07084  -0.78
##  Residual             0.393514 0.62731
## Number of obs: 5171, groups:  HHid, 1360
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)    5.62727    0.06207  90.664
## Year           0.01845    0.00333   5.542
## logmu          0.75084    0.01191  63.053
## logratio       0.30907    0.01020  30.292
## URUrban        0.03301    0.05961   0.554
## CIInland       0.04643    0.02150   2.160
## logmu:URUrban -0.03695    0.03560  -1.038
##
## Correlation of Fixed Effects:
##             (Intr) Year   logmu  lograt URUrbn CIInln
## Year        -0.025
## logmu       -0.452  0.005
## logratio    -0.921 -0.180  0.231
## URUrban     -0.039 -0.022  0.286 -0.043
## CIInland    -0.232 -0.017 -0.171  0.122 -0.170
## logm:URUrbn  0.043  0.030 -0.300  0.033 -0.783  0.037
## convergence code: 0
## Model failed to converge with max|grad| = 0.0854849 (tol = 0.002, component 1)
```

```
int2<-lmer(Y~Year+logmu+logratio+UR+CI+Year*logmu+(Year|HHid),data=workingdata,REML=FALSE)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.050971
## (tol = 0.002, component 1)
```

```r
summary(int2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Y ~ Year + logmu + logratio + UR + CI + Year * logmu + (Year |
##     HHid)
##    Data: workingdata
##
##      AIC      BIC   logLik deviance df.resid
##  10818.5  10890.5  -5398.2  10796.5     5160
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -13.7861  -0.4146   0.0436   0.4510   5.7858
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  HHid     (Intercept) 0.021372 0.14619
##           Year        0.004785 0.06917  -0.77
##  Residual             0.392523 0.62652
## Number of obs: 5171, groups:  HHid, 1360
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  5.729067   0.063906  89.648
## Year        -0.008409   0.005500  -1.529
## logmu        0.667015   0.017304  38.547
## logratio     0.311875   0.010175  30.651
## URUrban     -0.010203   0.037019  -0.276
## CIInland     0.049723   0.021431   2.320
## Year:logmu   0.019812   0.003239   6.117
##
## Correlation of Fixed Effects:
##            (Intr) Year   logmu  lograt URUrbn CIInln
## Year       -0.217
## logmu      -0.484  0.611
## logratio   -0.883 -0.139  0.137
## URUrban    -0.003 -0.018  0.039 -0.026
## CIInland   -0.221 -0.026 -0.124  0.122 -0.226
## Year:logmu  0.254 -0.800 -0.756  0.038  0.024  0.019
## convergence code: 0
## Model failed to converge with max|grad| = 0.050971 (tol = 0.002, component 1)
```

2. Variance-Covariance selection

1. Unstructured

```r
library(nlme)
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:lme4':
##
##     lmList
```

```
Ctr<-lmeControl(maxIter=200,msMaxIter=100,tolerance = 1e-2,msTol = 1e-3,returnObject=TRUE)
var1<-lme(Y~logmu+Year+logratio+UR+CI,
           random=~Year|HHid,correlation=corSymm(form=~1),data=workingdata,control =Ctr)
```

```
## Warning in lme.formula(Y ~ logmu + Year + logratio + UR + CI, random = ~Year | : nlminb problem, conv
##   message = iteration limit reached without convergence (10)
```

```
summary(var1)
```

```
## Linear mixed-effects model fit by REML
##  Data: workingdata
##        AIC      BIC    logLik
##   10620.79 10751.79 -5290.397
##
## Random effects:
##  Formula: ~Year | HHid
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.1404425 (Intr)
## Year        0.0756627 -0.998
## Residual    0.5976228
##
## Correlation Structure: General
##  Formula: ~1 | HHid
##  Parameter estimate(s):
##  Correlation:
##   1      2      3      4
## 2  0.155
## 3  0.174  0.028
## 4  0.274  0.016 -0.195
## 5  0.336  0.017 -0.355 -0.586
## Fixed effects: Y ~ logmu + Year + logratio + UR + CI
##                 Value  Std.Error   DF  t-value p-value
## (Intercept)  5.595561 0.06066768 3808 92.23297  0.0000
## logmu        0.751138 0.01133107 3808 66.29010  0.0000
## Year         0.020106 0.00308063 3808  6.52652  0.0000
## logratio     0.313604 0.00995732 3808 31.49482  0.0000
## URUrban     -0.016446 0.03773999 1357 -0.43576  0.6631
## CIInland     0.050194 0.02179393 1357  2.30313  0.0214
##  Correlation:
##          (Intr) logmu  Year   lograt URUrbn
## logmu    -0.464
## Year     -0.001  0.007
## logratio -0.920  0.253 -0.207
## URUrban  -0.010  0.086  0.007 -0.028
## CIInland -0.234 -0.167 -0.016  0.114 -0.227
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
```

```
## -14.34564336  -0.45784212   0.04257211   0.48160672   6.05147799
##
## Number of Observations: 5171
## Number of Groups: 1360
```

2. Compound symmetry

```
var2<-lme(Y~logmu+Year+logratio+UR+CI,
          random=~Year|HHid,correlation=corCompSymm(form=~1),data=workingdata)
summary(var2)
```

```
## Linear mixed-effects model fit by REML
##  Data: workingdata
##        AIC      BIC    logLik
##   10897.77 10969.81 -5437.884
##
## Random effects:
##  Formula: ~Year | HHid
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.18660187 (Intr)
## Year        0.07091433 -0.663
## Residual    0.62011357
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | HHid
##  Parameter estimate(s):
##         Rho
## -0.02402036
## Fixed effects: Y ~ logmu + Year + logratio + UR + CI
##                 Value  Std.Error   DF  t-value p-value
## (Intercept)  5.630158 0.06206202 3808 90.71825  0.0000
## logmu        0.747169 0.01137511 3808 65.68451  0.0000
## Year         0.018566 0.00333098 3808  5.57370  0.0000
## logratio     0.309391 0.01020495 3808 30.31776  0.0000
## URUrban     -0.015411 0.03715452 1357 -0.41479  0.6784
## CIInland     0.047254 0.02151207 1357  2.19661  0.0282
##  Correlation:
##          (Intr) logmu  Year   lograt URUrbn
## logmu    -0.461
## Year     -0.026  0.015
## logratio -0.923  0.253 -0.181
## URUrban  -0.009  0.087  0.002 -0.027
## CIInland -0.233 -0.168 -0.019  0.121 -0.227
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -13.76487297  -0.41444149   0.04524897   0.44308141   5.64947514
##
## Number of Observations: 5171
## Number of Groups: 1360
```

3. AR1

```
var3<-lme(Y~logmu+Year+logratio+UR+CI,
          random=~Year|HHid,correlation=corAR1(form=~1),data=workingdata,control =Ctr)
summary(var3)
```

```
## Linear mixed-effects model fit by REML
##  Data: workingdata
##        AIC      BIC    logLik
##   10896.53 10968.58 -5437.267
##
## Random effects:
##  Formula: ~Year | HHid
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.19599154 (Intr)
## Year        0.07218814 -0.698
## Residual    0.61978143
##
## Correlation Structure: AR(1)
##  Formula: ~1 | HHid
##  Parameter estimate(s):
##       Phi
## -0.033231
## Fixed effects: Y ~ logmu + Year + logratio + UR + CI
##                 Value  Std.Error   DF  t-value p-value
## (Intercept)  5.629300 0.06205410 3808 90.71600  0.0000
## logmu        0.747622 0.01136684 3808 65.77222  0.0000
## Year         0.018650 0.00331851 3808  5.62010  0.0000
## logratio     0.309328 0.01020853 3808 30.30094  0.0000
## URUrban     -0.015423 0.03711750 1357 -0.41551  0.6778
## CIInland     0.047056 0.02150498 1357  2.18813  0.0288
##  Correlation:
##          (Intr) logmu  Year   lograt URUrbn
## logmu    -0.460
## Year     -0.025  0.015
## logratio -0.924  0.253 -0.183
## URUrban  -0.009  0.087  0.002 -0.028
## CIInland -0.234 -0.168 -0.019  0.121 -0.226
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -13.73789260  -0.41383242   0.04318411   0.44659298   5.68707741
##
## Number of Observations: 5171
## Number of Groups: 1360
```

4. Exp

```
var4<-lme(Y~logmu+Year+logratio+UR+CI,
          random=~Year|HHid,correlation=corExp(form=~1),data=workingdata,control =Ctr)
summary(var4)
```

```
## Linear mixed-effects model fit by REML
##  Data: workingdata
##        AIC      BIC    logLik
##   10897.77 10969.81 -5437.884
##
## Random effects:
##  Formula: ~Year | HHid
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.15994685 (Intr)
## Year        0.07091432 -0.773
## Residual    0.62751724
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~1 | HHid
##  Parameter estimate(s):
##     range
## 0.0591006
## Fixed effects: Y ~ logmu + Year + logratio + UR + CI
##                 Value  Std.Error   DF  t-value p-value
## (Intercept)  5.630158 0.06206201 3808 90.71827  0.0000
## logmu        0.747169 0.01137511 3808 65.68455  0.0000
## Year         0.018566 0.00333098 3808  5.57370  0.0000
## logratio     0.309391 0.01020495 3808 30.31777  0.0000
## URUrban     -0.015411 0.03715449 1357 -0.41479  0.6784
## CIInland     0.047254 0.02151205 1357  2.19662  0.0282
##  Correlation:
##          (Intr) logmu  Year   lograt URUrbn
## logmu    -0.461
## Year     -0.026  0.015
## logratio -0.923  0.253 -0.181
## URUrban  -0.009  0.087  0.002 -0.027
## CIInland -0.233 -0.168 -0.019  0.121 -0.227
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -13.7762585  -0.4166378   0.0425202   0.4531121   5.7571794
##
## Number of Observations: 5171
## Number of Groups: 1360
```

5. CAR

```
var5<-lme(Y~logmu+Year+logratio+UR+CI,
          random=~Year|HHid,correlation=corCAR1(form=~1),data=workingdata,control =Ctr)
summary(var5)
```

```
## Linear mixed-effects model fit by REML
##  Data: workingdata
##        AIC      BIC    logLik
##   10897.77 10969.81 -5437.884
##
## Random effects:
##  Formula: ~Year | HHid
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.15996453 (Intr)
## Year        0.07091541 -0.773
## Residual    0.62751595
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~1 | HHid
##  Parameter estimate(s):
##          Phi
## 7.231454e-08
## Fixed effects: Y ~ logmu + Year + logratio + UR + CI
##                 Value  Std.Error   DF  t-value p-value
## (Intercept)  5.630158 0.06206202 3808 90.71824  0.0000
## logmu        0.747169 0.01137511 3808 65.68459  0.0000
## Year         0.018566 0.00333100 3808  5.57369  0.0000
## logratio     0.309391 0.01020494 3808 30.31776  0.0000
## URUrban     -0.015411 0.03715450 1357 -0.41478  0.6784
## CIInland     0.047254 0.02151206 1357  2.19662  0.0282
##  Correlation:
##          (Intr) logmu  Year   lograt URUrbn
## logmu    -0.461
## Year     -0.027  0.015
## logratio -0.923  0.253 -0.181
## URUrban  -0.009  0.087  0.002 -0.027
## CIInland -0.233 -0.168 -0.019  0.121 -0.227
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -13.77629352  -0.41663820   0.04251749   0.45311455   5.75719498
##
## Number of Observations: 5171
## Number of Groups: 1360
```

Variance of Y

```
D=matrix(c(0.019724,-0.010605,-0.010605,0.005725),nrow=2)
Z=matrix(c(1,0,1,2,1,5,1,7,1,11),nrow=2)
sigmasq=0.5976228^2
R<-matrix(c(1,0.155,0.174,0.274,0.336,0.155,1,0.028,0.016,0.017,0.174,0.028,1,-0.195,-0.355,0.274,0.016
R<-sigmasq*R
t(Z)%*%D%*%Z+R
```

```
##            [,1]       [,2]       [,3]       [,4]       [,5]
## [1,] 0.37687701 0.05387272 0.02884362 0.04334893 0.03378700
```

```
## [2,] 0.05387272 0.35735701 0.01273928 0.01014345 0.01388060
## [3,] 0.02884362 0.01273928 0.41395201 0.02319416 0.03812968
## [4,] 0.04334893 0.01014345 0.02319416 0.50893201 0.06036734
## [5,] 0.02307241 0.01388060 0.03812968 0.06036734 0.83629201
```