To install the package from a local machine use the code below. Note that the packages *lme4* and *stringr* are also necessary and need to be installed.

```
#install.packages("modSel_0.1.0.tar.gz", repos = NULL, type = "source")
library(modSel)
#?modelSelection
```

Example 1

We will illustrate the use of the model selection algorithm with *Boston* data set that can be found in the *MASS* library.

```
library(MASS)
head(Boston)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12
##   lstat medv
## 1  4.98 24.0
## 2  9.14 21.6
## 3  4.03 34.7
## 4  2.94 33.4
## 5  5.33 36.2
## 6  5.21 28.7
```

For the sake of illustration, we will use the effect of variable *chas* as a random effect. The fixed effects we would like to consider are listed in the object *fxd* below.

```
# make chas a factor to use it as a random effect
Boston$chas = as.factor(Boston$chas)
rndm  = c("(1 | chas)")
fxd = c("tax", "indus", "nox", "rm", "ptratio", "black", "medv",
          "I(tax^2)", "I(indus^2)", "I(nox^2)", "I(ptratio^2)",
          "I(tax^3)", "I(tax^4)",
          "tax:indus", "tax:nox", "tax:rm", "tax:ptratio",
          "tax:black", "tax:medv",
          "indus:nox", "indus:rm", "indus:ptratio", "rm:ptratio",
          "tax:nox:indus", "tax:rm:ptratio")
```

We would also like to keep the main effect of variable *rm* in our models, so we will include it in all models.

```
include = c("rm")
```

Suppose that we are interested in the variable *crim*, per capita crime rate by town, and we would like to model the crime rate with the mixed model

$$
\begin{aligned}
crim \quad \sim \quad & (1 \mid chas) + tax + indus + nox + ptratio + black + medv \\
& + I(tax^2) + I(indus^2) + I(nox^2) + I(ptratio^2) \\
& + I(tax^3) + I(tax^4) \\
& + tax : indus + tax : nox + tax : rm + tax : ptratio + tax : black + tax : medv \\
& + indus : nox + indus : rm + indus : ptratio + rm : ptratio \\
& + tax : nox : indus + tax : rm : ptratio
\end{aligned} \tag{1}
$$

To find the most parsimonious model, we will start with the full model (1) and will use *backward* selection to drop the terms from the full model. We will use *BIC* as a discriminating criterion between different models. The input *sinkit* creates a file "results.txt" in *R* working directory and prints the output to this file that would, otherwise, be printed to the *R* console.

```
# mixed model/backward/BIC
mod1 <- modelSelection(response = "crim", random = rndm, fixed = fxd, keep = include,
               dat = Boston, direction = "backward", method = "BIC",
               sinkit = "results.txt")
mod1
```

```
## $fullModel
## [1] "lmer(crim  ~  (1 | chas) + tax + indus + nox + rm + ptratio + black + medv + I(tax^2) + I(indus
##
## $finalModel
## [1] "lmer(crim  ~  (1 | chas) + tax + rm + medv + I(tax^2) + I(tax^3) + tax:medv + I(tax^4))"
##
## $termsKept
## [1] "rm"
##
## $direction
## [1] "backward"
##
## $method
## [1] "BIC"
##
## $origMethodValue
## [1] 3446.42
##
## $path
##      Term dropped      BIC
## 1        indus:rm 3440.239
## 2   tax:indus:nox 3434.081
## 3       tax:black 3427.966
## 4    I(ptratio^2) 3422.193
## 5   indus:ptratio 3416.994
## 6  tax:rm:ptratio 3412.479
## 7      rm:ptratio 3406.811
## 8     tax:ptratio 3401.415
## 9          tax:rm 3396.509
## 10        ptratio 3391.821
## 11     I(indus^2) 3388.289
## 12        tax:nox 3383.944
## 13      indus:nox 3380.222
## 14       I(nox^2) 3375.064
```

```
## 15            nox 3368.840
## 16      tax:indus 3365.964
## 17          indus 3360.037
## 18          black 3357.858
## 19       I(tax^4) 3381.847
##
## $runningTime
## [1] "11.69 secs"
```

The best model obtained through backward selection using *BIC* is given by

$$crim \sim (1 \mid chas) + tax + rm + medv + I(tax^2) + I(tax^3) + tax : medv + I(tax^4) \quad (2)$$

The BIC values for models (1) and (2) were 3446.42 and 3381.847, respectively. Model (2) was obtained in 11.69 secs seconds.

Here is a partial output that is stored in "results.txt" that informs the user about the selection process.

```
[1] "------------------------------------------------------------"
[1] "------------------------------------------------------------"
The model now has BIC value of 3446.422.

The model now has 26 terms, of which 25 are fixed effects and 1 are/is random effect(s).

There are 10 terms that are eligible to be dropped and these are:

 [1] "tax:black"      "tax:medv"       "indus:rm"       "indus:ptratio"
 [5] "tax:indus:nox"  "tax:rm:ptratio" "I(indus^2)"     "I(nox^2)"
 [9] "I(ptratio^2)"   "I(tax^4)"


R is going to fit 10 model(s), each time removing one variable that is eligible to be dropped from the c

The 10 models were fitted in 2.03 secs.

The top five alternative models with the lowest BIC are:

   Term Dropped      AIC      BIC    logLik deviance LRT p-Value
6      indus:rm 3326.123 3440.239 -1636.061 3272.123   0.8341859
8 tax:indus:nox 3326.147 3440.263 -1636.073 3272.147   0.7946782
7     tax:black 3326.212 3440.328 -1636.106 3272.212   0.7157352
3  I(ptratio^2) 3326.613 3440.729 -1636.306 3272.613   0.4651557
5 indus:ptratio 3327.127 3441.243 -1636.563 3273.127   0.3060152


The term that will be dropped is: indus:rm.

The updated model now has BIC value of 3440.239.

The updated model is now:
crim ~ (1 | chas) + tax + indus + nox + rm + ptratio + black +
    medv + I(tax^2) + I(indus^2) + I(nox^2) + I(ptratio^2) +
    I(tax^3) + I(tax^4) + tax:indus + tax:nox + tax:rm + tax:ptratio +
    tax:black + tax:medv + indus:nox + indus:ptratio + rm:ptratio +
```

```
    tax:indus:nox + tax:rm:ptratio
<environment: 0x107963f58>


[1] "------------------------------------------------------------"
[1] "------------------------------------------------------------"
The model now has BIC value of 3440.239.

The model now has 25 terms, of which 24 are fixed effects and 1 are/is random effect(s).

There are 9 terms that are eligible to be dropped and these are:

[1] "tax:black"      "tax:medv"       "indus:ptratio"  "tax:indus:nox"
[5] "tax:rm:ptratio" "I(indus^2)"     "I(nox^2)"       "I(ptratio^2)"
[9] "I(tax^4)"


R is going to fit 9 model(s), each time removing one variable that is eligible to be dropped from the o

The 9 models were fitted in 0.54 secs.

The top five alternative models with the lowest BIC are:

   Term Dropped      AIC      BIC    logLik deviance LRT p-Value
7 tax:indus:nox 3324.191 3434.081 -1636.096 3272.191   0.7939316
6     tax:black 3324.248 3434.138 -1636.124 3272.248   0.7233357
3  I(ptratio^2) 3324.647 3434.537 -1636.323 3272.647   0.4691850
5 indus:ptratio 3325.178 3435.068 -1636.589 3273.178   0.3044141
1    I(indus^2) 3325.903 3435.792 -1636.951 3273.903   0.1822032


The term that will be dropped is: tax:indus:nox.

The updated model now has BIC value of 3434.081.

The updated model is now:
crim ~ (1 | chas) + tax + indus + nox + rm + ptratio + black +
    medv + I(tax^2) + I(indus^2) + I(nox^2) + I(ptratio^2) +
    I(tax^3) + I(tax^4) + tax:indus + tax:nox + tax:rm + tax:ptratio +
    tax:black + tax:medv + indus:nox + indus:ptratio + rm:ptratio +
    tax:rm:ptratio
<environment: 0x107963f58>
```

Example 2

This example uses the same data set, but it builds the final model using *forward* selection, where the models are compared with the likelihood ratio test with $\alpha = 0.05$. The procedure's output is saved in *results2.txt* file and is not shown here.

```
# mixed model/forward/LRT/alpha default
mod2 <- modelSelection(response = "crim", random = rndm, fixed = fxd,
                direction = "forward", keep = include, method = "LRT", dat = Boston,
                sinkit = "results2.txt")
```

Here is are the results of the forward selection procedure:

```
mod2
```

```
## $finalModel
## [1] "lmer(crim  ~  (1 | chas) + rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4) + black)
##
## $termsKept
## [1] "rm"
##
## $direction
## [1] "forward"
##
## $method
## [1] "LRT"
##
## $path
##                                                                          LRT pValue
## c("tax:medv", "tax", "medv")                                        7.692058e-18
## c("I(tax^2)", "I(tax^3)", "I(tax^4)")                               1.320418e-02
## black                                                               4.425196e-02
## c("tax:rm:ptratio", "tax:rm", "tax:ptratio", "rm:ptratio", "ptratio") 9.060073e-01
##
## $alpha
## [1] 0.05
##
## $runningTime
## [1] "4.83 secs"
```

The output shows that, to the model containing $crim \sim (1 \mid chas)$, in round one the algorithm added *tax:medv, tax, medv*. Then, in round two, to the model $crim \sim (1 \mid chas) + tax : medv + tax + medv$, the algorithm added *I(tax^2), I(tax^3), I(tax^4)*. In round three, the algorithm added variable *black*. In round four, the three way interaction *tax:rm:ptratio* could not be added since the LRT p-value $\approx 0.91 > 0.05$. The final model was then given by

$$crim \sim (1 \mid chas) + black + rm + tax + medv + tax : medv + I(tax^2) + I(tax^3) + I(tax^4).$$

Example 3

This example uses the same data set and it builds the final model using *forward* selection. Since there aren't any random effects and we are potentially adding more than one term at a time (if an interaction is considered for inclusion, it is included with its main effects), the models are compared with an F test and $\alpha = 0.025$. The procedure's output is not saved but is directly printed to the console.

```
# fixed model/forward/alpha = 0.025
mod3 <- modelSelection(response = "crim", fixed = fxd,
               direction = "forward", keep = include, dat = Boston, alpha = 0.025)
```

```
## [1] "------------------------------------------------------------"
## [1] "------------------------------------------------------------"
## The model now has 2 terms, including the intercept, of which 1 are/is fixed effect(s).
##
## The groups of terms that are eligible to be added are:
```

```
##
##    AddGroup
## 1  "tax:indus, tax, indus"
## 2  "tax:nox, tax, nox"
## 3  "tax:rm, tax"
## 4  "tax:ptratio, tax, ptratio"
## 5  "tax:black, tax, black"
## 6  "tax:medv, tax, medv"
## 7  "indus:nox, indus, nox"
## 8  "indus:rm, indus"
## 9  "indus:ptratio, indus, ptratio"
## 10 "rm:ptratio, ptratio"
## 11 "tax:indus:nox, tax:indus, tax:nox, indus:nox, tax, indus, nox"
## 12 "tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, tax, ptratio"
## 13 "tax, I(tax^2)"
## 14 "indus, I(indus^2)"
## 15 "nox, I(nox^2)"
## 16 "ptratio, I(ptratio^2)"
## 17 "tax, I(tax^2), I(tax^3)"
## 18 "tax, I(tax^2), I(tax^3), I(tax^4)"
## 19 "tax"
## 20 "indus"
## 21 "nox"
## 22 "ptratio"
## 23 "black"
## 24 "medv"
##
## The top five most significant group of terms are:
##
##                      Terms Considered F-test p-Value
## 6                tax:medv, tax, medv   2.923040e-54
## 18 tax, I(tax^2), I(tax^3), I(tax^4)   4.513501e-49
## 13                      tax, I(tax^2)   5.834653e-46
## 17           tax, I(tax^2), I(tax^3)   3.925719e-45
## 5              tax:black, tax, black   1.348082e-44
##
## The group of terms that will be added is: tax:medv + tax + medv.
##
## The updated model is now:
## crim ~ rm + tax:medv + tax + medv
## <environment: 0x7fd3681f67c0>
##
##
## [1] "----------------------------------------------------------------"
## [1] "----------------------------------------------------------------"
## The model now has 5 terms, including the intercept, of which 4 are/is fixed effect(s).
##
## The groups of terms that are eligible to be added are:
##
##    AddGroup
## 1  "tax:indus, indus"
## 2  "tax:nox, nox"
## 3  "tax:rm"
## 4  "tax:ptratio, ptratio"
```

```
## 5  "tax:black, black"
## 6  "indus:nox, indus, nox"
## 7  "indus:rm, indus"
## 8  "indus:ptratio, indus, ptratio"
## 9  "rm:ptratio, ptratio"
## 10 "tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox"
## 11 "tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, ptratio"
## 12 "I(tax^2)"
## 13 "indus, I(indus^2)"
## 14 "nox, I(nox^2)"
## 15 "ptratio, I(ptratio^2)"
## 16 "I(tax^2), I(tax^3)"
## 17 "I(tax^2), I(tax^3), I(tax^4)"
## 18 "indus"
## 19 "nox"
## 20 "ptratio"
## 21 "black"
##
## The top five most significant group of terms are:
##
##                                          Terms Considered F-test p-Value
## 17                             I(tax^2), I(tax^3), I(tax^4)   3.184184e-08
## 10 tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox   3.259510e-04
## 12                                                I(tax^2)   9.077710e-03
## 13                                       indus, I(indus^2)   9.684604e-03
## 21                                                   black   1.380694e-02
##
## The group of terms that will be added is: I(tax^2) + I(tax^3) + I(tax^4).
##
## The updated model is now:
## crim ~ rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4)
## <environment: 0x7fd3681f67c0>
##
##
## [1] "------------------------------------------------------------"
## [1] "------------------------------------------------------------"
## The model now has 8 terms, including the intercept, of which 7 are/is fixed effect(s).
##
## The groups of terms that are eligible to be added are:
##
##    AddGroup
## 1  "tax:indus, indus"
## 2  "tax:nox, nox"
## 3  "tax:rm"
## 4  "tax:ptratio, ptratio"
## 5  "tax:black, black"
## 6  "indus:nox, indus, nox"
## 7  "indus:rm, indus"
## 8  "indus:ptratio, indus, ptratio"
## 9  "rm:ptratio, ptratio"
## 10 "tax:indus:nox, tax:indus, tax:nox, indus:nox, indus, nox"
## 11 "tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, ptratio"
## 12 "indus, I(indus^2)"
## 13 "nox, I(nox^2)"
```

```
## 14 "ptratio, I(ptratio^2)"
## 15 "indus"
## 16 "nox"
## 17 "ptratio"
## 18 "black"
##
## The top five most significant group of terms are:
##
##                                              Terms Considered F-test p-Value
## 18                                                      black     0.04629174
## 11 tax:rm:ptratio, tax:rm, tax:ptratio, rm:ptratio, ptratio     0.08952398
## 4                                      tax:ptratio, ptratio     0.09601093
## 5                                          tax:black, black     0.13732558
## 1                                          tax:indus, indus     0.14690211
##
## The group of terms that will be added is: black.
##
## The updated model is now:
## crim ~ rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4) +
##     black
## <environment: 0x7fd3681f67c0>
##
##
## Adding 'black' will result in an F-test p-Value bigger than alpha, so the term(s) will not be added a
```

Here are the results of this forward selection procedure:

```
mod3
```

```
## $finalModel
## [1] "lm(crim  ~  rm + tax:medv + tax + medv + I(tax^2) + I(tax^3) + I(tax^4))"
##
## $termsKept
## [1] "rm"
##
## $direction
## [1] "forward"
##
## $path
##                      TermsAdded       pValue
## 1          tax:medv, tax, medv 2.923040e-54
## 2 I(tax^2), I(tax^3), I(tax^4) 3.184184e-08
## 3                         black 4.629174e-02
##
## $alpha
## [1] 0.025
##
## $runningTime
## [1] "0.47 secs"
```