R Notebook

```
suppressMessages(library(AER))
suppressMessages(library(sampleSelection))
data(tobin)
data(Mroz87)

tobin$durable_binary <- as.numeric(tobin$durable > 0)
Mroz87$kids <- as.numeric((Mroz87$kids5 + Mroz87$kids618) > 0)
```

```
model.tobit  <- tobit(durable ~ 1 + age + quant, data=tobin)
model.probit <- glm(durable_binary ~ 1 + age + quant, data=tobin, family=binomial(link = "probit"), x=T)
tobin$IMR    <- invMillsRatio(model.probit)$IMR1
model.heckit <- lm(durable ~ 1 + age + quant + IMR, data=tobin, subset=(durable>0))

summary(model.tobit)
```

```
##
## Call:
## tobit(formula = durable ~ 1 + age + quant, data = tobin)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##             20             13              7              0
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 15.14487   16.07945   0.942    0.346
## age         -0.12906    0.21858  -0.590    0.555
## quant       -0.04554    0.05825  -0.782    0.434
## Log(scale)   1.71785    0.31032   5.536  3.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Scale: 5.573
##
## Gaussian distribution
## Number of Newton-Raphson Iterations: 3
## Log-likelihood: -28.94 on 4 Df
## Wald-statistic: 1.124 on 2 Df, p-value: 0.57002
```

```
summary(model.probit)
```

```
##
## Call:
## glm(formula = durable_binary ~ 1 + age + quant, family = binomial(link = "probit"),
##     data = tobin, x = T)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2431  -0.9382  -0.7184   1.3989   1.5451
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  1.2929682  3.2342150   0.400    0.689
## age         -0.0344696  0.0398456  -0.865    0.387
## quant       -0.0001906  0.0115077  -0.017    0.987
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25.898  on 19  degrees of freedom
## Residual deviance: 25.169  on 17  degrees of freedom
## AIC: 31.169
##
## Number of Fisher Scoring iterations: 4
```

```
summary(model.heckit)
```

```
##
## Call:
## lm(formula = durable ~ 1 + age + quant + IMR, data = tobin, subset = (durable >
##     0))
##
## Residuals:
##       2       8      10      11      15      18      19
## -1.1340 -1.7857  0.1661  2.0472  0.1189  1.8061 -1.2186
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36.5481    50.5780   0.723    0.522
## age          -5.5133    12.8279  -0.430    0.696
## quant        -0.1629     0.1321  -1.234    0.305
## IMR         253.4042   555.3742   0.456    0.679
##
## Residual standard error: 2.118 on 3 degrees of freedom
## Multiple R-squared:  0.7889, Adjusted R-squared:  0.5778
## F-statistic: 3.737 on 3 and 3 DF,  p-value: 0.1538
```

Zur Kontrolle, dass die benutzten Algorithmen mit unserer Theorie übereinstimmen:

```
# ist invMillsRatio() wirklich "lambda(-x_i' * alpha_PROB)"
tobin$manualIMR <- dnorm(-model.probit$x %*% model.probit$coefficients) / (1 - pnorm(-model.probit$x %*%
abs(tobin$IMR - t(tobin$manualIMR)) < 10^-15
```

```
##         1    2    3    4    5    6    7    8    9   10   11   12   13   14
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##        15   16   17   18   19   20
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE
```

Nochmals mit einem anderen Datenset

```
model2.tobit  <- tobit(wage ~ 1 + exper + I(exper^2) + educ + city, data=Mroz87)
model2.probit <- glm(lfp ~ 1 + age + I(age^2) + faminc + kids + educ, data=Mroz87, family=binomial(link
Mroz87$IMR   <- invMillsRatio(model2.probit)$IMR1
model2.heckit <- lm(wage ~ 1 + exper + I(exper^2) + educ + city + IMR, data=Mroz87, subset=(lfp==1))

summary(model2.tobit)
```

```
##
## Call:
## tobit(formula = wage ~ 1 + exper + I(exper^2) + educ + city,
```

```
##     data = Mroz87)
##
## Observations:
##          Total  Left-censored     Uncensored Right-censored
##            753            325            428              0
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.394546   1.095091  -9.492  < 2e-16 ***
## exper         0.460704   0.068107   6.764 1.34e-11 ***
## I(exper^2)   -0.009064   0.002119  -4.278 1.89e-05 ***
## educ          0.641702   0.081177   7.905 2.68e-15 ***
## city         -0.087014   0.378424  -0.230    0.818
## Log(scale)    1.496149   0.036933  40.510  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Scale: 4.464
##
## Gaussian distribution
## Number of Newton-Raphson Iterations: 4
## Log-likelihood: -1463 on 6 Df
## Wald-statistic: 153.9 on 4 Df, p-value: < 2.22e-16
```

```
summary(model2.probit)
```

```
##
## Call:
## glm(formula = lfp ~ 1 + age + I(age^2) + faminc + kids + educ,
##     family = binomial(link = "probit"), data = Mroz87, x = T)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9205  -1.2261   0.7877   1.0634   1.6515
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.157e+00  1.404e+00  -2.961 0.003070 **
## age          1.854e-01  6.621e-02   2.800 0.005107 **
## I(age^2)    -2.426e-03  7.762e-04  -3.125 0.001775 **
## faminc       4.580e-06  4.306e-06   1.064 0.287417
## kids        -4.490e-01  1.300e-01  -3.453 0.000554 ***
## educ         9.818e-02  2.289e-02   4.289  1.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1029.7  on 752  degrees of freedom
## Residual deviance:  981.7  on 747  degrees of freedom
## AIC: 993.7
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(model2.heckit)
```

```
##
## Call:
## lm(formula = wage ~ 1 + exper + I(exper^2) + educ + city + IMR,
##     data = Mroz87, subset = (lfp == 1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6897 -1.6009 -0.4961  0.8481 21.2041
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.9712472  2.0387141  -0.476    0.634
## exper        0.0210613  0.0629715   0.334    0.738
## I(exper^2)   0.0001371  0.0018934   0.072    0.942
## educ         0.4170193  0.0990025   4.212 3.09e-05 ***
## city         0.4438385  0.3179214   1.396    0.163
## IMR         -1.0975873  1.2529008  -0.876    0.382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.112 on 422 degrees of freedom
## Multiple R-squared:  0.1264, Adjusted R-squared:  0.116
## F-statistic: 12.21 on 5 and 422 DF,  p-value: 4.576e-11
```