6.7.a. The data can be generated in EViews by means of the following program.

```
create exer6_7 u 1 200         ' create workfile
genr x = @trend(0)             ' generate series x, ystar and y
genr ystar = -10 + 0.1*x + nrnd
genr y = (ystar >= 0)
```

b. The theoretical odds ratios and the odds ratios in the sample can be computed as follows.

```
vector(5) tx            ' stores the 5 values of x
vector(5) tratio_b      ' theoretical odds ratios
vector(5) ratio_b       ' odds ratios in sample

tx.fill 60, 80, 100, 120, 140

for !i = 1 to 5         ' compute odds ratios

  tratio_b(!i) = @cnorm(-10+0.1*tx(!i))/(1-@cnorm(-10+0.1*tx(!i)))

  !s1 = (35+!i*20)     ' begin of sample
  !s2 = (45+!i*20)     ' end of sample
  smpl  !s1 !s2        ' adjust sample
  if @mean(y) <> 1 then
     ratio_b(!i) = @mean(y)/(1-@mean(y))
  else
     ratio_b(!i) = na
  endif

next                    ' end of for loop

smpl 1 200
scat x ystar            ' scatter diagram
```

The theoretical and sample odds ratios are shown in Table S 6.2. The sample odds ratio for $95 \leq x_i \leq 105$ (1.2) is quite close to the theoretical value of 1 for $x = 100$. For $55 \leq x_i \leq 65$ and $75 \leq x_i \leq 85$ all observed values of y_i are zero, so the sample odds ratio is also zero. Note that both intervals contain 11 observations, so that these outcomes could be expected on the basis of the theoretical odds ratios (of zero and 0.023 respectively). Further, the sample odds ratios for the intervals $115 \leq x_i \leq 125$ and $135 \leq x_i \leq 145$ can not be determined as all y_i have the value 1 in these two subsamples (as could be expected from the large theoretical odds ratios for $x = 120$ and $x = 140$).

Figure S 6.3 shows the scatter diagram of y^* against x. Clearly, around $x = 60$ and $x = 80$ there are no values $y^* > 0$, so that $y = 0$ in these intervals. For instance, for $x = 80$ there holds $y^* = -10 + 8 + \varepsilon_i$ so that $P[y^* > 0] = P[\varepsilon_i > 2] = 0.023$. For similar reasons, around $x = 120$ and $x = 140$ there are no values $y^* < 0$ (although in our simulation for $x_i = 127$ the corresponding value $y_i^* = 0.479$ comes quite close to zero), so that $y = 1$ in these intervals.

c. The results of OLS are shown in Table S 6.3, and the corresponding estimated odds ratios are given in the third column of Table S 6.4. The OLS estimates of the odds ratios are very imprecise. For small values of x the odds ratio is much overestimated, whereas for large values of x it is much underestimated. That is, the estimated effect of x on the odds ratio is much smaller than the actual effect.

theoretical		sample	
x	ratio	x_i	ratio
60	0.000	55-65	0.000
80	0.023	75-85	0.000
100	1.000	95-105	1.200
120	42.956	115-125	na
140	31573.386	135-145	na

Table S 6.2: Part (b) : Theoretical and sample odds ratios

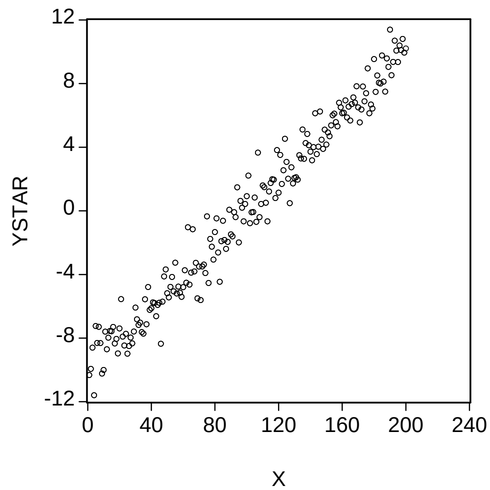

Figure S 6.3: Part (b) : Scatter diagram of y^* against x

Dependent Variable: Y
Method: Least Squares
Date: 11/11/02 Time: 15:10
Sample: 1 200
Included observations: 200

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.244573	0.036941	-6.620718	0.0000
X	0.007409	0.000319	23.24516	0.0000

R-squared	0.731830	Mean dependent var	0.500000
Adjusted R-squared	0.730476	S.D. dependent var	0.501255
S.E. of regression	0.260230	Akaike info criterion	0.155449
Sum squared resid	13.40850	Schwarz criterion	0.188432
Log likelihood	-13.54488	F-statistic	540.3374
Durbin-Watson stat	1.118403	Prob(F-statistic)	0.000000

Table S 6.3: Part (c) : OLS estimates

x	Theoretical	OLS	Probit	Logit
60	0.000	0.250	0.000	0.001
80	0.023	0.534	0.026	0.034
100	1.000	0.985	0.919	0.921
120	42.956	1.813	29.722	24.513
140	31573.386	3.822	10955.660	657.636

Table S 6.4: Part (c), (d) and (f) : Estimated odds ratios of OLS, Probit and Logit

d. The estimated probit model is shown in Table S 6.5. The estimated coefficients are close to the parameters of the DGP, and the corresponding odds ratios in the fourth column of Table S 6.4 are relatively close to the theoretical values. This is as expected, because the probit model assumes the normal distribution for the unobserved error terms and this is the correct specification for our simulated data.

Dependent Variable: Y
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 11/11/02 Time: 15:10
Sample: 1 200
Included observations: 200
Convergence achieved after 8 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-9.540121	2.058268	-4.635023	0.0000
X	0.094873	0.020323	4.668147	0.0000

Mean dependent var	0.500000	S.D. dependent var	0.501255
S.E. of regression	0.183079	Akaike info criterion	0.215760
Sum squared resid	6.636514	Schwarz criterion	0.248744
Log likelihood	-19.57604	Hannan-Quinn criter.	0.229108
Restr. log likelihood	-138.6294	Avg. log likelihood	-0.097880
LR statistic (1 df)	238.1068	McFadden R-squared	0.858789
Probability(LR stat)	0.000000		
Obs with Dep=0	100	Total obs	200
Obs with Dep=1	100		

Table S 6.5: Part (d) : Estimated probit model

e. The estimated logit model is shown in Table S 6.6. The estimated coefficients are approximately a factor 1.6 larger than the probit estimates in Table S 6.5. This would also be expected from the results in Section 6.1.2 concerning the relation between the parameters of logit and probit models.

Dependent Variable: Y
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 11/11/02 Time: 15:10
Sample: 1 200
Included observations: 200
Convergence achieved after 8 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-16.50910	3.711253	-4.448390	0.0000
X	0.164270	0.036707	4.475157	0.0000

Mean dependent var	0.500000	S.D. dependent var	0.501255
S.E. of regression	0.184227	Akaike info criterion	0.220272
Sum squared resid	6.720027	Schwarz criterion	0.253255
Log likelihood	-20.02722	Hannan-Quinn criter.	0.233620
Restr. log likelihood	-138.6294	Avg. log likelihood	-0.100136
LR statistic (1 df)	237.2044	McFadden R-squared	0.855534
Probability(LR stat)	0.000000		
Obs with Dep=0	100	Total obs	200
Obs with Dep=1	100		

Table S 6.6: Part (e) : Estimated logit model

f. The odds ratios of the logit model of part (e) are shown in the last column of Table S 6.4. The logit model performs quite well as the odds ratios are relatively close to the theoretical values. Note that the logit model is not the correct model specification, as the DGP corresponds to a probit model. However, notwithstanding this specification error, the logit model is well able to estimate the involved odds and the results are comparable to the ones of the (correctly specified) probit model of part (d).