2. Binary response models

2.1 Linear probability model

Binary dependent variables in a microeconometric analysis:

These qualitative variables have exactly two possible categories and thus take two values, namely one and zero

Examples for microeconometric analyses with binary response models:

- Analysis of the factors that explain whether a person is employed or unemployed
- Analysis of the factors that explain whether a person uses a specific means of transportation or other means of transportation (as multinomial variable)
- Analysis of the factors that explain whether a person strongly agrees with a statement (based on an ordinal scale) or not
- Analysis of the factors that explain whether a household owns a certain insurance or not
- Analysis of the factors that explain whether the profits of a firm are at least as high as a specific amount or are lower than this amount
- Analysis of the factors that explain whether a firm has realized an innovation in the last three years or not

If y_i is a binary dependent variable with $y_i = 1$ or $y_i = 0$, $x_i = (x_{i1},\ldots, x_{ik})'$ is a vector of k explanatory variables (including a constant), and $\beta = (\beta_1,\ldots, \beta_k)'$ is the corresponding k-dimensional parameter vector, a microeconometric model can simply be specified as a multiple linear regression model (for i = 1,…, n):

$$y_i = \beta' x_i + \varepsilon_i$$

Such a linear regression model with a binary dependent variable is called linear probability model. With $E(\varepsilon_i|x_i) = 0$ it follows:

$$E(y_i|x_i) = \beta' x_i$$

Since y_i is a binary variable with $y_i = 1$ or $y_i = 0$, it is Bernoulli distributed with parameter p_i and the following probability function:

$$f_i(y_i; p_i) = p_i^{y_i}(1-p_i)^{1-y_i} \quad \text{for } y_i = 0, 1$$

In the linear probability model it follows:

$$p_i = p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = E(y_i|x_i) = \beta' x_i$$

Interpretation of the slope parameters in the linear probability model:

- Due to the binary character of y_i, the slope parameters β_h (h = 2,…, k) cannot be interpreted as the change in y_i for a one-unit increase of the explanatory variable x_{ih} as in common linear regression models
- Instead, β_h (h = 2,…, k) indicates the change in the probability $p_i(x_i, \beta)$ that y_i takes the value one if x_{ih} increases by one unit (for a quantitative explanatory variable), ceteris paribus

If all other explanatory variables are held fixed, this means formally:

$$\Delta p_i(x_i, \beta) = \Delta P\left(y_i = 1|x_i, \beta\right) = \beta_h \Delta x_{ih}$$

In line with the OLS estimation of the parameters in common linear regression models, the unknown regression parameters $\beta_1,\ldots, \beta_k$ in the linear probability model can also be estimated by OLS. This leads to the OLS estimator of the parameter vector $\hat{\beta} = (\hat{\beta}_1,\ldots, \hat{\beta}_k)$. It follows:

- The estimator of the dependent variable is $\hat{y}_i = \hat{\beta}'x_i$, which is the estimator $\hat{p}_i(x_i, \beta)$ of the probability that y_i takes the value one
- The estimator of the slope parameter $\hat{\beta}_h$ (h = 2,…, k) indicates the change in the estimated probability $\hat{p}_i(x_i, \beta)$ if x_{ih} increases by one unit (for a quantitative explanatory variable), ceteris paribus

Problem:

Since y_i is Bernoulli distributed with parameter $p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = \beta'x_i$ and $\varepsilon_i = -\beta'x_i$ for $y_i = 0$ and $\varepsilon_i = 1 - \beta'x_i$ for $y_i = 1$, it follows for the conditional variance of y_i and the conditional variance of the error term ε_i:

$$\mathrm{Var}(y_i|x_i) = \mathrm{Var}(\varepsilon_i|x_i) = \beta'x_i(1-\beta'x_i)$$

The conditional variance of the error term can therefore necessarily not be constant, but depends on the explanatory variables which leads to heteroskedasticity. As a consequence, either an alternative estimation method instead of the OLS method or heteroskedasticity robust t statistics should at least be used.

Example: Determinants of labor force participation of married women (I)

By using a linear probability model, the effect of other sources of income (in 1000 dollar) including the earnings of the husband (nwifeinc), the years of education (educ), the years of labor market experience (exper), the squared years of labor market experience (expersq), the age in years (age), the number of children less than six years old (kidslt6), and the number of children between six and 18 years of age (kidsge6) on the labor force participation of married women is examined. The dependent dummy variable inlf takes the value one if the woman is employed. The following OLS regression equation was estimated with n = 753 women which also reports heteroskedasticity robustly estimated standard deviations of the estimated parameters (in brackets) in addition to conventionally estimated standard deviations ($R^2 = 0.264$):

$$\widehat{inlf} = \underset{\substack{(0.154)\\ [0.152]}}{0.586} - \underset{\substack{(0.001)\\ [0.002]}}{0.003}nwifeinc + \underset{\substack{(0.007)\\ [0.007]}}{0.038}educ + \underset{\substack{(0.006)\\ [0.006]}}{0.039}exper - \underset{\substack{(0.000)\\ [0.000]}}{0.001}expersq$$

$$- \underset{\substack{(0.002)\\ [0.002]}}{0.016}age - \underset{\substack{(0.034)\\ [0.032]}}{0.262}kidslt6 + \underset{\substack{(0.013)\\ [0.014]}}{0.013}kidsge6$$

Example: Determinants of labor force participation of married women (II)

```
reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust

Linear regression                                      Number of obs =     753
                                                       F(  7,   745) =   62.48
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2642
                                                       Root MSE      =  .42713

------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0034052   .0015249    -2.23   0.026    -.0063988   -.0004115
        educ |   .0379953    .007266     5.23   0.000      .023731    .0522596
       exper |   .0394924     .00581     6.80   0.000     .0280864    .0508983
     expersq |  -.0005963     .00019    -3.14   0.002    -.0009693   -.0002233
         age |  -.0160908    .002399    -6.71   0.000    -.0208004   -.0113812
     kidslt6 |  -.2618105   .0317832    -8.24   0.000    -.3242058   -.1994152
     kidsge6 |   .0130122   .0135329     0.96   0.337     -.013555    .0395795
       _cons |   .5855192   .1522599     3.85   0.000     .2866098    .8844287
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (III)

Interpretation:

- Based on both types of t statistics, all explanatory variables except kidsge6 have significant effects
- One more year of education leads to an estimated increase of the probability of labor force participation by 0.038 or 3.8 percentage points (ceteris paribus)
- Ten more years of education therefore imply an estimated increase of the probability of labor force participation by $0.038 \cdot 10 = 0.38$ or 38 percentage points, which is a large effect
- An increase of nwifeinc by 10000 dollars (i.e. $\Delta nwifeinc = 10$) leads to an estimated decrease of the probability of labor force participation by 0.034 or 3.4 percentage points, which is not a large effect
- The effect of exper depends on the value of exper: An increase of labor market experience by one year leads to an estimated change of the probability of labor force participation by $0.039 - 2 \cdot 0.0006 \cdot exper = 0.039 - 0.0012 \cdot exper$
- One additional child less than six years old implies an estimated decrease of the probability of labor force participation by 0.262 or 26.2 percentage points, which is a huge effect

Evaluation of the use of linear probability models:

- In line with the OLS estimation in linear regression models including hetero-skedastic error terms, the parameters are easy to estimate and the estima-ted slope parameters are easy to interpret since they are partial effects
- However: The estimated probabilities $\hat{p}_i(x_i, \beta) = \widehat{P}(y_i = 1|x_i, \beta)$ that the depen-dent variable y_i takes the value one are not restricted to the interval between zero and one, i.e. for specific combinations of values for the explanatory va-riables, the estimated probabilities can be smaller than zero or greater than one which is not possible by definition
- It is possible that a probability is not linearly related to an explanatory variab-le for all possible values: For example, the previous microeconometric ana-lysis implies an estimated decrease of the probability of labor force participa-tion by 0.262 for the increase from zero children less than six years to one child. This decrease is equally estimated for the increase from one child to two children, although it seems more realistic that the decrease of the pro-bability is stronger for the increase from zero children to one child. In fact, the previous analysis implies that four additional children lead to an estima-ted decrease of the probability of labor force participation by $0.262 \cdot 4 = 1.048$ which is clearly not possible by definition.

→ As a consequence, the linear probability model is not used frequently in practice

2.2 Binary probit and logit models

Binary dependent variables y_i in a microeconometric model with the vector of k explanatory variables $x_i = (x_{i1},\dots, x_{ik})'$ and the corresponding k-dimensional parameter vector $\beta = (\beta_1,\dots, \beta_k)'$ are generally Bernoulli distributed with the following probability function (for i = 1,…, n):

$$f_i\left(y_i; x_i, \beta\right) = p_i(x_i, \beta)^{y_i}\left[1-p_i(x_i, \beta)\right]^{1-y_i}$$

$$= P(y_i = 1|x_i, \beta)^{y_i}\left[1-P(y_i = 1|x_i, \beta)\right]^{1-y_i} \quad \text{for } y_i = 0, 1$$

Different binary response models result from different specifications of the probability $p_i(x_i, \beta) = P(y_i = 1|x_i)$ that y_i takes the value one. In the case of linear probability models, this probability is identical to $\beta'x$ which does not ensure that the value is between zero and one as discussed above.

Such values can be guaranteed by several nonlinear functions $F_i(x_i, \beta) = F_i(\beta'x_i)$ and particularly by distribution functions of arbitrary random variables. In the case of binary probit models, $F_i(\beta'x_i) = \Phi_i(\beta'x_i)$ is the value of the distribution function of the standard normal distribution at the linear function $\beta'x_i$:

$$F_i(\beta'x_i) = \Phi_i(\beta'x_i) = p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = \int_{-\infty}^{\beta'x_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

The probabilities $p_i(x_i, \beta)$ in binary probit models are therefore calculated by integration.

In the case of binary logit models, $F_i(\beta' x_i) = \Lambda_i(\beta' x_i)$ is the value of the distribution function of the standard logistic distribution at the linear function $\beta' x_i$:

$$F_i(\beta' x_i) = \Lambda_i(\beta' x_i) = p_i(x_i, \beta) = P(y_i = 1 | x_i, \beta) = \frac{e^{\beta' x_i}}{1+e^{\beta' x_i}}$$

In contrast to binary probit models, the probabilities $p_i(x_i, \beta)$ in binary logit models need not be calculated by integration, but can be derived from a closed form.

→ Despite the substantial differences in the functional forms, the probabilities $p_i(x_i, \beta) = P(y_i = 1|x_i, \beta)$ in binary probit and logit models are very similar (except for a constant scaling factor, see below) so that the choice between them makes little difference in practice (in contrast to the difference between multinomial probit and logit models, see later)

Binary probit and logit models can also be motivated by an underlying continuous latent variable y_i^* (which can be interpreted as a utility, see later) which depends on $\beta' x_i$ and an error term ε_i (for i = 1,…, n):

$$y_i^* = \beta' x_i + \varepsilon_i$$

If the latent variables would be observable, this would lead to linear regression models. However, latent variables are not observable. But they can be related to the observed binary dependent variables y_i:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

It follows for the probability $p_i(x_i, \beta)$ that y_i takes the value one:

$$p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = P(y_i^* \geq 0|x_i, \beta) = P(\beta' x_i + \varepsilon_i \geq 0) = P(\varepsilon_i \geq -\beta' x_i)$$

Different binary response models can be derived by different distribution assumptions for ε_i.

If ε_i has a standard normal distribution (with expected value of zero and variance one), this leads to binary probit models. If ε_i has a standard logistic distribution with expected value of zero and variance $\pi^2/3$, this leads to binary logit models. In both cases ε_i is symmetrically distributed around zero so that it follows in binary probit and logit models:

$$p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = \Phi_i(\beta' x_i) = 1 - \Phi_i(-\beta' x_i)$$

$$p_i(x_i, \beta) = P(y_i = 1|x_i, \beta) = \Lambda_i(\beta' x_i) = 1 - \Lambda_i(-\beta' x_i)$$

The assumption of a known variance of ε_i is not problematic since this variance is not identified in both binary response models and thus cannot be estimated besides β so that it has to be normalized. The assumption of zero as the threshold as discussed above is also unproblematic if a constant is included in $\beta' x_i$ (therefore, such a constant should generally be incorporated).

Interpretation of a parameter β_h in binary response models with respect to the (partial) effect of the corresponding explanatory variable x_{ih} ($h = 2,\ldots, k$) on the probability $p_i(x_i, \beta) = P(y_i = 1|x_i, \beta)$:

- The parameter β_h cannot be interpreted as simply as in the linear probability model, i.e. it cannot be interpreted as the change in $p_i(x_i, \beta)$ if x_{ih} increases by one unit (for a quantitative explanatory variable), ceteris paribus
- Instead, the (partial) marginal probability effects of x_{ih} in binary response models are as follows (for $i = 1,\ldots, n$):

$$\frac{\partial p_i(x_i, \beta)}{\partial x_{ih}} = \frac{\partial F_i(\beta' x_i)}{\partial x_{ih}} = \frac{dF_i(\beta' x_i)}{d(\beta' x_i)} \frac{\partial \beta' x_i}{\partial x_{ih}} = f_i(\beta' x_i)\beta_h$$

While $F_i(\beta' x_i)$ is the distribution function of the standard normal distribution in binary probit models and the distribution function of the standard logistic distribution in binary logit models, $f_i(\beta' x_i)$ is the density function of the standard normal distribution in binary probit models and the density function of the standard logistic distribution in binary logit models.

- If all other explanatory variables are held fixed, it thus follows for a change Δx_{ih}:

$$\Delta p_i(x_i, \beta) \approx \left[f(\beta' x_i)\beta_h\right]\Delta x_{ih}$$

The smaller the change Δx_{ih}, the better is this linear approximation.

(Partial) marginal probability effects of x_{ih} in binary probit models with $\varphi_i(\beta' x_i)$ as the density function of the standard normal distribution:

$$\frac{\partial p_i(x_i, \beta)}{\partial x_{ih}} = \frac{\partial \Phi_i(\beta' x_i)}{\partial x_{ih}} = \frac{d\Phi_i(\beta' x_i)}{d(\beta' x_i)} \frac{\partial \beta' x_i}{\partial x_{ih}} = \varphi_i(\beta' x_i)\beta_h$$

(Partial) marginal probability effects of x_{ih} in binary logit models with $\Lambda_i(\beta' x_i)$ as the distribution function of the standard logistic distribution :

$$\frac{\partial p_i(x_i, \beta)}{\partial x_{ih}} = \frac{\partial \Lambda_i(\beta' x_i)}{\partial x_{ih}} = \frac{d\Lambda_i(\beta' x_i)}{d(\beta' x_i)} \frac{\partial \beta' x_i}{\partial x_{ih}}$$

$$= \frac{e^{\beta' x_i}}{\left(1+e^{\beta' x_i}\right)^2}\beta_h = \Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right]\beta_h$$

Important aspects of (partial) marginal probability effects in binary probit and logit models:

- The sign of the parameter β_h indicates the direction of the marginal probability effects
- The marginal probability effects are maximal for $\beta' x_i = 0$ since the density functions are maximal at this value
- The marginal probability effects vary with different values not only of the explanatory variable x_{ih} but also with different values of all other explanatory variables and thus across different observations

→ The formulas for the (partial) marginal probability effects in binary response models show that the relative effects of two explanatory variables x_{ih} and x_{ig} do not depend on $f_i(\beta'x_i)$ and therefore not on the explanatory variables in x_i since the ratio of the marginal probability effects of x_{ih} and x_{ig} is:

$$\frac{f_i(\beta' x_i)\beta_h}{f_i(\beta' x_i)\beta_g} = \frac{\beta_h}{\beta_g}$$

In practice it is generally interesting to evaluate the marginal probability effect of the explanatory variable x_{ih} for a typical observation i (e.g. person, household, firm). With $x_i = (x_{i1},\ldots, x_{ik})'$ as vector of k explanatory variables for i a possible calculation is the average (partial) marginal probability effect ($AMPE_h$) of x_{ih} which includes the marginal probability effects $f_i(\beta'x_i)\beta_h$ for each observation i = 1,…, n:

$$AMPE_h = \frac{1}{n}\sum_{i=1}^{n} f_i(\beta' x_i)\beta_h$$

$$AMPE_h = \frac{1}{n}\sum_{i=1}^{n} \varphi_i(\beta' x_i)\beta_h \quad \text{in the binary probit model}$$

$$AMPE_h = \frac{1}{n}\sum_{i=1}^{n} \Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right]\beta_h \quad \text{in the binary logit model}$$

Another possible computation refers to the (partial) marginal probability effect ($MPEM_{ih}$) of x_{ih} that is calculated at the mean $\bar{x} = 1/n\sum_{i=1}^{n} x_i$ of the explanatory variables across the observations i = 1,…, n:

$$MPEM_{ih} = f_i(\beta'\bar{x})\beta_h$$

$$MPEM_{ih} = \varphi_i(\beta'\bar{x})\beta_h \quad \text{in the binary probit model}$$

$$MPEM_{ih} = \Lambda_i(\beta'\bar{x})\left[1\text{-}\Lambda(\beta'\bar{x})\right]\beta_h \quad \text{in the binary logit model}$$

Due to the non-linearity of the approaches, the two types of effects are generally not identical, although they can be very similar in practice.

The discussion so far has implicitly assumed continuous explanatory variables. However, if explanatory variables are discrete (e.g. numbers of children) or qualitative (e.g. gender) or if larger changes of continuous explanatory variables are considered, the computation of the (partial) effect of an infinitesimal change of an explanatory variable x_{ih} can be very inaccurate. Therefore, a discrete change of $p_i(x_i, \beta)$ due to a discrete change Δx_{ih} is (for i = 1,…, n):

$$\Delta p_i(x_i, \beta) = F_i(\beta' x_i + \beta_h \Delta x_{ih}) - F_i(\beta' x_i)$$

$$\Delta p_i(x_i, \beta) = \Phi_i(\beta' x_i + \beta_h \Delta x_{ih}) - \Phi_i(\beta' x_i) \quad \text{in the binary probit model}$$

$$\Delta p_i(x_i, \beta) = \Lambda_i(\beta' x_i + \beta_h \Delta x_{ih}) - \Lambda_i(\beta' x_i) \quad \text{in the binary logit model}$$

Again it is possible to calculate (partial) average (discrete) effects and (discrete) effects at the mean $\bar{x} = 1/n\sum_{i=1}^{n} x_i$ of the explanatory variables (see later).

2.3 ML estimation and testing in binary probit and logit models

As discussed above, the dependent variables y_i in binary response models are dummy variables and thus are generally Bernoulli distributed with the probability $p_i(x_i, \beta)$. Based on a random sample (x_i, y_i) for i = 1,…, n observations, the log-likelihood function therefore is:

$$\begin{aligned}\log L(\beta) &= \left[y_1 \log p_1(x_1, \beta) + (1-y_1)\log\left(1-p_1(x_1, \beta)\right)\right] + \cdots + \\ &\quad \left[y_n \log p_n(x_n, \beta) + (1-y_n)\log\left(1-p_n(x_n, \beta)\right)\right] \\ &= \sum_{i=1}^{n}\left[y_i \log p_i(x_i, \beta) + (1-y_i)\log\left(1-p_i(x_i, \beta)\right)\right] \\ &= \sum_{i=1}^{n}\left[y_i \log F_i(\beta' x_i) + (1-y_i)\log\left(1-F_i(\beta' x_i)\right)\right]\end{aligned}$$

In binary probit and logit models it follows:

$$\log L(\beta) = \sum_{i=1}^{n}\left[y_i \log \Phi_i(\beta' x_i) + (1-y_i)\log\left(1-\Phi_i(\beta' x_i)\right)\right]$$

$$\log L(\beta) = \sum_{i=1}^{n}\left[y_i \log \Lambda_i(\beta' x_i) + (1-y_i)\log\left(1-\Lambda_i(\beta' x_i)\right)\right]$$

The score has the following form:

$$\frac{\partial \log L(\beta)}{\partial \beta} = s(\beta) = \sum_{i=1}^{n} \frac{y_i - p_i(x_i, \beta)}{p_i(x_i, \beta)\left(1-p_i(x_i, \beta)\right)} \frac{\partial p_i(x_i, \beta)}{\partial \beta}$$

$$= \sum_{i=1}^{n} \frac{y_i - F_i(\beta' x_i)}{F_i(\beta' x_i)\left[1-F_i(\beta' x_i)\right]} f_i(\beta' x_i) x_i$$

In binary probit and logit models it follows:

$$s(\beta) = \sum_{i=1}^{n} \frac{y_i - \Phi_i(\beta' x_i)}{\Phi_i(\beta' x_i)\left[1-\Phi_i(\beta' x_i)\right]} \varphi_i(\beta' x_i) x_i$$

$$s(\beta) = \sum_{i=1}^{n} \frac{y_i - \Lambda_i(\beta' x_i)}{\Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right]} \Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right] x_i$$

$$= \sum_{i=1}^{n} \left[y_i - \Lambda_i(\beta' x_i)\right] x_i$$

The ML estimator $\hat{\beta}$ solves the first-order conditions for maximizing the log-likelihood function, i.e. s(β) = 0. In general, a closed-form solution for $\hat{\beta}$ is not available so that iterative numerical optimization algorithms must be applied. Furthermore, it can be shown that the log-likelihood functions in both binary probit and logit models are globally concave so that the optimization process leads to a unique (global) maximum (and not a minimum).

Also the ML estimator $\hat{\beta}$ in binary response models is approximately normally distributed for large but finite samples of n observations with the following variance covariance matrix:

$$\mathrm{Var}(\hat{\beta}) = \left[nI(\beta)\right]^{-1} = -E\left[nH_i(\beta)\right]^{-1} = -E\left[H(\beta)\right]^{-1} = \left[\sum_{i=1}^{n} \frac{f_i(\beta' x_i)^2 x_i x_i'}{F_i(\beta' x_i)\left[1-F_i(\beta' x_i)\right]}\right]^{-1}$$

It follows for large samples:

$$\hat{\beta} \overset{appr}{\sim} N\left(\beta; -E\left[H(\beta)\right]^{-1}\right) \quad \text{or} \quad \hat{\beta} \overset{appr}{\sim} N\left(\beta; \left[\sum_{i=1}^{n} \frac{f_i(\beta' x_i)^2 x_i x_i'}{F_i(\beta' x_i)\left[1-F_i(\beta' x_i)\right]}\right]^{-1}\right)$$

$$\hat{\beta} \overset{appr}{\sim} N\left(\beta; \left[\sum_{i=1}^{n} \frac{\varphi_i(\beta' x_i)^2 x_i x_i'}{\Phi_i(\beta' x_i)\left[1-\Phi_i(\beta' x_i)\right]}\right]^{-1}\right) \quad \text{in the binary probit model}$$

$$\hat{\beta} \overset{appr}{\sim} N\left(\beta; \left[\sum_{i=1}^{n} \frac{\left(\Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right]\right)^2 x_i x_i'}{\Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right]}\right]^{-1}\right)$$

$$\overset{appr}{\sim} N\left(\beta; \left[\sum_{i=1}^{n} \Lambda_i(\beta' x_i)\left[1-\Lambda_i(\beta' x_i)\right] x_i x_i'\right]^{-1}\right) \quad \text{in the binary logit model}$$

The variance covariance matrix of the ML estimator $\hat{\beta}$ (just as the corresponding information matrix) in binary response models for large samples is generally unknown and thus has to be estimated for the construction of confidence intervals and particularly for statistical tests. The standard approach for this estimator includes the corresponding Hessian matrix at $\hat{\beta}$ for all observations:

$$\hat{Var}(\hat{\beta}) = \left(\sum_{i=1}^{n} \frac{f_i(\hat{\beta}' x_i)^2 x_i x_i'}{F_i(\hat{\beta}' x_i)\left[1-F_i(\hat{\beta}' x_i)\right]} \right)^{-1}$$

$$\hat{Var}(\hat{\beta}) = \left[\sum_{i=1}^{n} \frac{\varphi_i(\hat{\beta}' x_i)^2 x_i x_i'}{\Phi_i(\hat{\beta}' x_i)\left[1-\Phi_i(\hat{\beta}' x_i)\right]} \right]^{-1} \quad \text{in the binary probit model}$$

$$\hat{Var}(\hat{\beta}) = \left[\sum_{i=1}^{n} \Lambda_i(\hat{\beta}' x_i)\left[1-\Lambda_i(\hat{\beta}' x_i)\right] x_i x_i' \right]^{-1} \quad \text{in the binary logit model}$$

For the estimation of the variance covariance matrix of $\hat{\beta}$, it is generally also possible to exclusively include the score at the ML estimator $\hat{\beta}$ or both the Hessian matrix and the score to receive a robust version in the case of a model misspecification. The most important elements of the estimated variance covariance matrix of $\hat{\beta}$ are on the diagonal and indicate the estimated variances of the ML estimated parameters.

Remarks for empirical studies with binary probit and logit models:

- The square roots of the diagonal elements of the estimated variance covariance matrixes represent the estimated standard deviations of the ML estimated parameters and are reported by econometric software packages such as STATA by default. These values are the basis for the construction of confidence intervals and the calculation of z statistics.
- The estimated variance covariance matrixes are the basis for general Wald test statistics (score tests are not often considered in binary probit and logit models)
- Econometric software packages such as STATA additionally report the maximum values of the log-likelihood functions by default that are always negative in binary probit and logit models since the natural logarithms of the corresponding probabilities (values between zero and one) are negative
- These maximum values of the log-likelihood functions are the basis for likelihood ratio test statistics. The most common test statistic is reported by econometric software packages such as STATA by default and refers (in accordance to the F test in linear regression models) to the null hypothesis that none of the explanatory variables has an effect.
- The pseudo R^2 is also reported by econometric software packages such as STATA (whereas another popular goodness-of-fit measure in binary probit and logit models, namely the percentage correctly predicted, is not reported)

Example: Determinants of labor force participation of married women (I)

As in the previous example, the effect of other sources of income in 1000 dollar, the years of education, the simple and squared years of labor market experience, the age in years, the number of children less than six years old, and the number of children between six and 18 years of age on the labor force participation is examined with data from n = 753 married women. However, not a linear probability model, but binary probit and logit models are now used. The ML estimation of the binary probit model with STATA leads to the following results:

```
probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(7)      =     227.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.30219                       Pseudo R2       =     0.2206

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593     0.53   0.595    -.7267472    1.266901
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (II)

The corresponding ML estimation of the binary logit model with STATA leads to the following results:

```
logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Logistic regression                               Number of obs   =        753
                                                  LR chi2(7)      =     226.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.76515                       Pseudo R2       =     0.2197

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0213452   .0084214    -2.53   0.011    -.0378509   -.0048394
        educ |   .2211704   .0434396     5.09   0.000     .1360303    .3063105
       exper |   .2058695   .0320569     6.42   0.000     .1430391    .2686999
     expersq |  -.0031541   .0010161    -3.10   0.002    -.0051456   -.0011626
         age |  -.0880244    .014573    -6.04   0.000     -.116587   -.0594618
     kidslt6 |  -1.443354   .2035849    -7.09   0.000    -1.842373   -1.044335
     kidsge6 |   .0601122   .0747897     0.80   0.422     -.086473    .2066974
       _cons |   .4254524   .8603697     0.49   0.621    -1.260841    2.111746
------------------------------------------------------------------------------
```

An example for the presentation of all these estimation results (in linear probability, binary probit, and binary logit models) in empirical studies (which typically includes at least the parameter estimates, the z statistics or estimated standard deviations of the estimated parameters, and some information about the significance of the effect of the explanatory variables) is as follows:

Example: Determinants of labor force participation of married women (III)

ML estimates (z statistics), dependent variable: labor force participation (inlf)			
Explanatory variables	Linear probability model	Binary probit model	Binary logit model
nwifeinc	-0.003** (-2.23)	-0.012** (-2.48)	-0.021** (-2.53)
educ	0.038*** (5.23)	0.131*** (5.18)	0.221*** (5.09)
exper	0.039*** (6.80)	0.123*** (6.59)	0.206*** (6.42)
expersq	-0.001*** (-3.14)	-0.002*** (-3.15)	-0.003*** (-3.10)
age	-0.016*** (-6.71)	-0.053*** (-6.23)	-0.088*** (-6.04)
kidslt6	-0.262*** (-8.24)	-0.868*** (-7.33)	-1.443*** (-7.09)
kidsge6	0.013 (0.96)	0.036 (0.83)	0.060 (0.80)
constant	0.586 (3.85)	0.270 (0.53)	0.425 (0.49)

Note: *** (**, *) means that the appropriate explanatory variable has an effect at the 1% (5%, 10%) significance level, n = 753

Example: Determinants of labor force participation of married women (IV)

Interpretation:

- The estimation results in all three models are qualitatively very similar, i.e. the signs of the parameter estimates are identical and the same explanatory variables have a significant effect across the approaches
- However, the magnitudes of the estimated effects are not directly comparable on the basis of the parameter estimates due to the different estimators of average marginal probability effects (or marginal probability effects at the mean of the explanatory variables, see later)
- In empirical applications with binary probit and logit models (and possibly linear probability models) a quick way to compare the parameter estimates is based on the different scale factors in the marginal (probability) effects: A rough rule of thumb implies that the parameter estimates in binary probit models can be multiplied by 1.6 to make them comparable with the estimates in binary logit models (or conversely divided by 0.625) and divided by 2.5 to make them comparable with the estimates in linear probability models.

Example: Determinants of labor force participation of married women (V)

Wald and likelihood ratio tests:

As an example, the null hypothesis that neither kidslt6 nor kidsge6 has any effect on the labor force participation, i.e. that the two parameters of kidslt6 and kidsge6 are both zero, is tested in the binary probit model. The command for the Wald test in STATA is:

```
test kidslt6=kidsge6=0

 ( 1)  [inlf]kidslt6 - [inlf]kidsge6 = 0
 ( 2)  [inlf]kidslt6 = 0

           chi2(  2) =   56.70
         Prob > chi2 =    0.0000
```

The application of the likelihood ratio test requires both the unrestricted and restricted ML estimation. After the unrestricted ML estimation the following STATA command for saving the estimation results is necessary (the choice of the name “unrestricted” is arbitrary):

```
estimates store unrestricted
```

After the restricted ML estimation a corresponding STATA command for saving the respective estimation results is necessary (the choice of the name “restricted” is again arbitrary).

Example: Determinants of labor force participation of married women (VI)

```
probit inlf nwifeinc educ exper expersq age

Probit regression                               Number of obs   =        753
                                                LR chi2(5)      =     164.13
                                                Prob > chi2     =     0.0000
Log likelihood = -432.80875                     Pseudo R2       =     0.1594

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0112811   .0046216    -2.44   0.015    -.0203393   -.0022229
        educ |   .1051111   .0239911     4.38   0.000     .0580894    .1521327
       exper |   .1253479   .0181857     6.89   0.000     .0897046    .1609912
     expersq |  -.0020406   .0005906    -3.46   0.001    -.0031982    -.000883
         age |  -.0286742    .006845    -4.19   0.000    -.0420902   -.0152581
       _cons |  -.6190137   .4165484    -1.49   0.137    -1.435434    .1974062
------------------------------------------------------------------------------

estimates store restricted
```

Command for the likelihood ratio test in STATA:

```
lrtest unrestricted restricted

Likelihood-ratio test                               LR chi2(2)  =     63.01
(Assumption: restricted nested in unrestricted)     Prob > chi2 =    0.0000
```

Both the Wald test and the likelihood ratio test therefore lead to the rejection of the null hypothesis at very low significance levels.

The estimator of the probabilities $p_i(x_i, \beta) = P(y_i = 1|x_i, \beta)$ for i in binary probit and logit models is based on the ML estimator $\hat{\beta}$ and the invariance principle as discussed in section 1.2:

$$\hat{p}_i(x_i, \hat{\beta}) = \hat{P}(y_i = 1|x_i, \hat{\beta}) = \Phi_i(\hat{\beta}'x_i)$$

$$\hat{p}_i(x_i, \hat{\beta}) = \hat{P}(y_i = 1|x_i, \hat{\beta}) = \Lambda_i(\hat{\beta}'x_i)$$

It follows for the estimator of the (partial) marginal probability effects of an explanatory variable x_{ih} in binary probit and logit models:

$$\frac{\partial \hat{p}_i(x_i, \hat{\beta})}{\partial x_{ih}} = \varphi_i(\hat{\beta}'x_i)\hat{\beta}_h$$

$$\frac{\partial \hat{p}_i(x_i, \hat{\beta})}{\partial x_{ih}} = \Lambda_i(\hat{\beta}'x_i)\left[1 - \Lambda_i(\hat{\beta}'x_i)\right]\hat{\beta}_h$$

Based on $y_1,\ldots, y_n$ and $x_1,\ldots, x_n$, it follows for the estimator of the average probabilities $p_i(x_i, \beta)$ across all i = 1,…, n observations in binary probit and logit models:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_i(x_i, \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\Phi_i(\hat{\beta}'x_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_i(x_i, \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\Lambda_i(\hat{\beta}'x_i)$$

For the estimators of the probability $p_i(x_i, \beta)$ at the mean $\bar{x} = 1/n\sum_{i=1}^{n} x_i$ of the explanatory variables in binary probit and logit models it follows:

$$\hat{p}_i(\bar{x}, \hat{\beta}) = \Phi_i(\hat{\beta}'\bar{x})$$

$$\hat{p}_i(\bar{x}, \hat{\beta}) = \Lambda_i(\hat{\beta}'\bar{x})$$

Based on the estimators of the (partial) marginal probability effects of an explanatory variable x_{ih} for each i, the estimator of the average (partial) marginal probability effects of x_{ih} across all i and the corresponding estimator of the (partial) marginal probability effects of x_{ih} at the mean $\bar{x}$ of the explanatory variables in binary probit and logit models are as follows:

$$A\hat{M}PE_h = \frac{1}{n}\sum_{i=1}^{n} \varphi_i(\hat{\beta}' x_i)\hat{\beta}_h$$

$$A\hat{M}PE_h = \frac{1}{n}\sum_{i=1}^{n} \Lambda_i(\hat{\beta}' x_i)\left[1 - \Lambda_i(\hat{\beta}' x_i)\right]\hat{\beta}_h$$

$$M\hat{P}EM_{ih} = \varphi_i(\hat{\beta}'\bar{x})\hat{\beta}_h$$

$$M\hat{P}EM_{ih} = \Lambda_i(\hat{\beta}'\bar{x})\left[1 - \Lambda_i(\hat{\beta}'\bar{x})\right]\hat{\beta}_h$$

The variance of the estimated marginal probability effects, the variance of the estimated average marginal probability effects, and the variance of the estimated marginal probability effects at the mean $\bar{x}$ of the explanatory variables can be estimated by using the Delta method.

The estimators of a discrete change of $p_i(x_i, \beta)$ due to a discrete change Δx_{ih} of an explanatory variable x_{ih} in binary probit and logit models are as follows:

$$\Delta\hat{p}_i(x_i, \hat{\beta}) = \Phi_i(\hat{\beta}' x_i + \hat{\beta}_h \Delta x_{ih}) - \Phi_i(\hat{\beta}' x_i)$$

$$\Delta\hat{p}_i(x_i, \hat{\beta}) = \Lambda_i(\hat{\beta}' x_i + \hat{\beta}_h \Delta x_{ih}) - \Lambda_i(\hat{\beta}' x_i)$$

Based on these estimators for all i = 1,…, n observations, the estimator of average discrete changes (ADC_h) of $p_i(x_i, \beta)$ across all i due to a discrete change Δx_{ih} and the corresponding estimator of discrete changes (DCM_{ih}) of $p_i(x_i, \beta)$ at the means $\bar{x}_1,\dots,\bar{x}_{h-1},\bar{x}_{h+1},\dots,\bar{x}_k$ of the other explanatory variables across i in binary probit and logit models are as follows:

$$A\hat{D}C_h = \frac{1}{n}\sum_{i=1}^{n}\left[\Phi_i(\hat{\beta}' x_i + \hat{\beta}_h \Delta x_{ih}) - \Phi_i(\hat{\beta}' x_i)\right]$$

$$A\hat{D}C_h = \frac{1}{n}\sum_{i=1}^{n}\left[\Lambda_i(\hat{\beta}' x_i + \hat{\beta}_h \Delta x_{ih}) - \Lambda_i(\hat{\beta}' x_i)\right]$$

$$D\hat{C}M_{ih} = \Phi_i\left(\hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{h-1}\bar{x}_{h-1} + \hat{\beta}_h(x_{ih} + \Delta x_{ih}) + \hat{\beta}_{h+1}\bar{x}_{h+1} + \dots + \hat{\beta}_k\bar{x}_k\right)$$
$$- \Phi_i\left(\hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{h-1}\bar{x}_{h-1} + \hat{\beta}_h x_{ih} + \hat{\beta}_{h+1}\bar{x}_{h+1} + \dots + \hat{\beta}_k\bar{x}_k\right)$$

$$D\hat{C}M_{ih} = \Lambda_i\left(\hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{h-1}\bar{x}_{h-1} + \hat{\beta}_h(x_{ih} + \Delta x_{ih}) + \hat{\beta}_{h+1}\bar{x}_{h+1} + \dots + \hat{\beta}_k\bar{x}_k\right)$$
$$- \Lambda_i\left(\hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{h-1}\bar{x}_{h-1} + \hat{\beta}_h x_{ih} + \hat{\beta}_{h+1}\bar{x}_{h+1} + \dots + \hat{\beta}_k\bar{x}_k\right)$$

Example: Determinants of labor force participation of married women (I)

Similar to the previous example, the determinants of the labor force participation are examined with data from n = 753 married women. However, the squared years of labor market experience are now not included as explanatory variable. The ML estimation of the binary probit model with STATA leads to the following results:

```
probit inlf nwifeinc educ exper age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(6)      =     217.31
                                                  Prob > chi2     =     0.0000
Log likelihood = -406.21886                       Pseudo R2       =     0.2110

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0115648   .0047942    -2.41   0.016    -.0209613   -.0021684
        educ |   .1336902   .0251346     5.32   0.000     .0844273    .1829531
       exper |   .0702165    .007571     9.27   0.000     .0553775    .0850555
         age |  -.0555548   .0083447    -6.66   0.000    -.0719101   -.0391995
     kidslt6 |  -.8742923   .1175098    -7.44   0.000    -1.104607   -.6439773
     kidsge6 |   .0345459   .0429862     0.80   0.422    -.0497055    .1187974
       _cons |   .5795817    .496205     1.17   0.243    -.3929623    1.552126
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (II)

The estimation of the average probability of labor force participation across all women and the estimation of the probability at the means of the explanatory variables in the binary probit model with STATA leads to the following results:

```
margins

Predictive margins                               Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5690098   .0155159    36.67   0.000     .5385992    .5994204
------------------------------------------------------------------------------

margins, atmeans

Adjusted predictions                             Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : nwifeinc        =    20.12896 (mean)
               educ            =    12.28685 (mean)
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5837979   .0199027    29.33   0.000     .5447893    .6228064
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (III)

It is also possible to estimate probabilities for a specific value of an explanatory variable. For example, the estimation of the average probability of labor force participation for educ = 10 and the estimation of the probability at the means of all other explanatory variables with STATA leads to the following results:

```
margins, at (educ=10)

Predictive margins                              Number of obs     =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : educ            =          10

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4728826   .0238422    19.83   0.000     .4261527    .5196124
------------------------------------------------------------------------------


margins, at((means)_all educ=10)

Adjusted predictions                            Number of obs     =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : nwifeinc        =    20.12896 (mean)
               educ            =          10
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4625106   .0299787    15.43   0.000     .4037535    .5212677
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (IV)

The estimation of the average marginal probability effects with STATA leads to the following results:

```
margins, dydx(*)

Average marginal effects                          Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper age kidslt6 kidsge6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |   -.003532     .00145    -2.44   0.015    -.0063739   -.0006901
        educ |   .0408301   .0072785     5.61   0.000     .0265645    .0550958
       exper |   .0214447   .0019236    11.15   0.000     .0176745     .025215
         age |  -.0169669   .0023171    -7.32   0.000    -.0215084   -.0124254
     kidslt6 |  -.2670162   .0317994    -8.40   0.000    -.3293418   -.2046906
     kidsge6 |   .0105506   .0131118     0.80   0.421    -.0151481    .0362493
------------------------------------------------------------------------------
```

These results refer to infinitesimal changes of continuous explanatory variables. Due to the difference between marginal changes and discrete changes, the interpretation that e.g. one more year of education leads to an estimated increase of the probability of labor force participation in the amount of 0.04 (or about 4 percentage points) is indeed a more or less good approximation, but not the exact value. The smaller the change in the explanatory variable, the better is the linear approximation.

Example: Determinants of labor force participation of married women (V)

The corresponding estimation of the marginal probability effects at the means of the explanatory variables with STATA leads to the following results:

```
margins, dydx(*) atmeans

Conditional marginal effects                      Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper age kidslt6 kidsge6
at           : nwifeinc        =    20.12896 (mean)
               educ            =    12.28685 (mean)
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0045115   .0018701    -2.41   0.016    -.0081769   -.0008462
        educ |   .0521537   .0098013     5.32   0.000     .0329436    .0713638
       exper |   .0273921   .0029446     9.30   0.000     .0216207    .0331634
         age |  -.0216724   .0032551    -6.66   0.000    -.0280522   -.0152926
     kidslt6 |   -.341069   .0459363    -7.42   0.000    -.4311026   -.2510354
     kidsge6 |   .0134767   .0167726     0.80   0.422     -.019397    .0463503
------------------------------------------------------------------------------
```

For the case of discrete explanatory variables, however, these approximate estimates can be very inaccurate. In addition, discrete explanatory variables can only take specific values (e.g. zero and one in the binary case).

Example: Determinants of labor force participation of married women (VI)

Therefore, this analysis is often not very reasonable for discrete explanatory variables (especially in the binary case). This could be considered in STATA treatments of binary probit or logit models:

- First of all, the underlying STATA ML estimation has to make clear that an explanatory variable is discrete (i.e. that it is an integer) by prefixing "i." (this produces and includes the maximum number of dummy variables based on the underlying variable), e.g. "i.educ" instead of "educ" as before
- The STATA command "margins educ" then reports the estimated average probabilities of labor force participation for all values of educ, whereas the command "margins educ, atmeans" reports the estimated probabilities of inlf = 1 for all values of educ at the means of all other explanatory variables
- The STATA command "margins, dydx(educ)" reports the estimated average probability changes of an increase from the base level five to the other values of educ, whereas the command "margins, dydx(educ) atmeans" reports the corresponding estimated probability changes of an increase from five to the other values of educ at the means of all other explanatory variables
- By including i.educ (i.e. 12 dummy variables) in the underlying ML estimation, all following estimates of probabilities and probability changes differ from the case that educ is included in the underlying ML estimation

---

Example: Determinants of labor force participation of married women (VII)

```
probit inlf nwifeinc i.educ exper age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(17)     =     226.99
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.37753                       Pseudo R2       =     0.2204

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |    -.01314   .0048546    -2.71   0.007    -.0226547   -.0036252
             |
        educ |
          6  |    .193736   .9081473     0.21   0.831      -1.5862    1.973672
          7  |   .1027714   .9212293     0.11   0.911    -1.702805    1.908348
          8  |   .3288946   .7550302     0.44   0.663    -1.150937    1.808727
          9  |   .1882056   .7671357     0.25   0.806    -1.315353    1.691764
         10  |    .296301   .7414178     0.40   0.689    -1.156851    1.749453
         11  |   .4195364   .7438581     0.56   0.573    -1.038399    1.877471
         12  |   .5693924   .7183331     0.79   0.428    -.8385146    1.977299
         13  |   .9000266   .7495592     1.20   0.230    -.5690824    2.369136
         14  |   .9836039   .7465668     1.32   0.188    -.4796402    2.446848
         15  |   .6430126   .8033719     0.80   0.423    -.9315674    2.217593
         16  |    1.01244   .7428555     1.36   0.173    -.4435304     2.46841
         17  |   1.847283    .769403     2.40   0.016     .3392812    3.355285
             |
       exper |   .0707929   .0076408     9.27   0.000     .0558171    .0857686
         age |  -.0558834   .0084173    -6.64   0.000     -.072381   -.0393857
     kidslt6 |  -.8956434   .1197816    -7.48   0.000    -1.130411   -.6608758
     kidsge6 |   .0395377   .0433334     0.91   0.362    -.0453942    .1244697
       _cons |     1.6006   .8250494     1.94   0.052    -.0164667    3.217667
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (VIII)

```
margins educ

Predictive margins                              Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |
          5  |   .3637474   .2146549     1.69   0.090    -.0569685    .7844633
          6  |   .4231178   .1767365     2.39   0.017     .0767206     .769515
          7  |   .3949565   .1803446     2.19   0.029     .0414877    .7484254
          8  |   .4655862   .0785653     5.93   0.000      .311601    .6195714
          9  |   .4213939    .087903     4.79   0.000     .2491071    .5936807
         10  |   .4552969   .0637233     7.14   0.000     .3304015    .5801922
         11  |   .4942804   .0664211     7.44   0.000     .3640973    .6244634
         12  |   .5416587   .0221742    24.43   0.000     .4981981    .5851192
         13  |   .6430566   .0642656    10.01   0.000     .5170984    .7690148
         14  |   .6674272   .0601789    11.09   0.000     .5494787    .7853756
         15  |   .5647403   .1134054     4.98   0.000     .3424698    .7870109
         16  |    .675679   .0539922    12.51   0.000     .5698563    .7815018
         17  |   .8678404   .0450665    19.26   0.000     .7795116    .9561692
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (IX)

```
margins educ, atmeans

Adjusted predictions                              Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : nwifeinc        =    20.12896 (mean)
               5.educ          =    .0053121 (mean)
               6.educ          =    .0079681 (mean)
               7.educ          =    .0106242 (mean)
               8.educ          =    .0398406 (mean)
               9.educ          =    .0332005 (mean)
               10.educ         =    .0584329 (mean)
               11.educ         =    .0571049 (mean)
               12.educ         =    .5059761 (mean)
               13.educ         =    .0584329 (mean)
               14.educ         =    .0677291 (mean)
               15.educ         =    .0185923 (mean)
               16.educ         =    .0756972 (mean)
               17.educ         =     .061089 (mean)
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |
          5  |   .3271236   .2581293     1.27   0.205    -.1788006    .8330478
          6  |   .3996961   .2189889     1.83   0.068    -.0295142    .8289064
          7  |   .3650102   .2206099     1.65   0.098    -.0673772    .7973976
          8  |   .4526475    .098629     4.59   0.000     .2593381    .6459569
          9  |   .3975614   .1090079     3.65   0.000     .1839099     .611213
         10  |   .4397635   .0798217     5.51   0.000     .2833159    .5962112
         11  |   .4886981   .0836898     5.84   0.000     .3246691    .6527272
         12  |   .5483614   .0279464    19.62   0.000     .4935876    .6031353
         13  |    .674422   .0785015     8.59   0.000      .520562     .828282
         14  |   .7039288     .07251     9.71   0.000     .5618118    .8460459
         15  |   .5773594   .1421988     4.06   0.000     .2986549    .8560639
         16  |   .7138168   .0644014    11.08   0.000     .5875925    .8400412
         17  |   .9191555   .0404382    22.73   0.000     .8398981    .9984129
------------------------------------------------------------------------------
```

Example: Determinants of labor force participation of married women (X)

It is again also possible to estimate probabilities that the dependent variable takes the value one for one specific value of an explanatory variable. For example, the estimation of the average probability of labor force participation for the case that educ = 10 with STATA leads to the following results (see also the results on page 36):

```
margins, at (educ=10)

Predictive margins                                  Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : educ            =          10

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4552969   .0637233     7.14   0.000     .3304015    .5801922
------------------------------------------------------------------------------
```

Remark: The command “margins, at (educ=10)” on page 31, which is based on an ML estimation that includes “educ” as explanatory variable and not “i.educ” and thus 12 education dummy variables leads to the estimated probability 0.4729 or 47.29%.

Example: Determinants of labor force participation of married women (XI)

The corresponding estimation of the probability that inlf = 1 for educ = 10 and at the means of all other explanatory variables with STATA leads to the following results (see also the results on page 37):

```
margins, at((means)_all educ=10)

Adjusted predictions                              Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
at           : nwifeinc        =    20.12896 (mean)
               educ            =          10
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4397635   .0798217     5.51   0.000     .2833159    .5962112
------------------------------------------------------------------------------
```

Remark: The command “margins, at((means)_all educ=10)” on page 31 on the basis of an ML estimation that includes “educ” as explanatory variable leads to the estimated probability 0.4625 or 46.25%.

Example: Determinants of labor force participation of married women (XII)

```
margins, dydx(educ)

Average marginal effects                          Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : 6.educ 7.educ 8.educ 9.educ 10.educ 11.educ 12.educ 13.educ 14.educ
15.educ 16.educ 17.educ

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |
          6  |   .0593703   .2766718     0.21   0.830    -.4828964    .6016371
          7  |   .0312091   .2789911     0.11   0.911    -.5156034    .5780216
          8  |   .1018388   .2278317     0.45   0.655    -.3447031    .5483807
          9  |   .0576465    .231372     0.25   0.803    -.3958343    .5111272
         10  |   .0915494   .2233221     0.41   0.682    -.3461538    .5292526
         11  |   .1305329   .2242491     0.58   0.561    -.3089872    .5700531
         12  |   .1779112   .2157371     0.82   0.410    -.2449256    .6007481
         13  |   .2793092   .2247299     1.24   0.214    -.1611534    .7197717
         14  |   .3036798   .2233503     1.36   0.174    -.1340788    .7414383
         15  |   .2009929   .2430107     0.83   0.408    -.2752993    .6772852
         16  |   .3119316   .2223107     1.40   0.161    -.1237893    .7476525
         17  |   .5040929   .2202834     2.29   0.022     .0723454    .9358405
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Example: Determinants of labor force participation of married women (XIII)

```
margins, dydx(educ) atmeans

Conditional marginal effects                    Number of obs   =        753
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : 6.educ 7.educ 8.educ 9.educ 10.educ 11.educ 12.educ 13.educ 14.educ 15.educ 16.educ 17.educ
at           : nwifeinc        =    20.12896 (mean)
               5.educ          =    .0053121 (mean)
               6.educ          =    .0079681 (mean)
               7.educ          =    .0106242 (mean)
               8.educ          =    .0398406 (mean)
               9.educ          =    .0332005 (mean)
               10.educ         =    .0584329 (mean)
               11.educ         =    .0571049 (mean)
               12.educ         =    .5059761 (mean)
               13.educ         =    .0584329 (mean)
               14.educ         =    .0677291 (mean)
               15.educ         =    .0185923 (mean)
               16.educ         =    .0756972 (mean)
               17.educ         =     .061089 (mean)
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |
          6  |   .0725725   .3367868     0.22   0.829    -.5875174    .7326624
          7  |   .0378866   .3380662     0.11   0.911    -.6247109    .7004841
          8  |   .1255239   .2754063     0.46   0.649    -.4142625    .6653103
          9  |   .0704378   .2794523     0.25   0.801    -.4772786    .6181542
         10  |   .1126399   .2694223     0.42   0.676     -.415418    .6406979
         11  |   .1615745     .27073     0.60   0.551    -.3690465    .6921956
         12  |   .2212378   .2594664     0.85   0.394    -.2873069    .7297825
         13  |   .3472984   .2704496     1.28   0.199     -.182773    .8773698
         14  |   .3768053   .2685153     1.40   0.161    -.1494751    .9030856
         15  |   .2502358   .2951319     0.85   0.397     -.328212    .8286837
         16  |   .3866933    .267043     1.45   0.148    -.1367013    .9100878
         17  |   .5920319    .262054     2.26   0.024     .0784155    1.105648
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

General remarks to binary response models:

- In the case of perfect prediction for a binary explanatory variable d_i, the ML estimation of binary probit and logit models is not possible. Perfect prediction for the dependent variable y_i arises if $y_i = 1$ whenever $d_i = 1$, if $y_i = 0$ whenever $d_i = 1$, if $y_i = 1$ whenever $d_i = 0$, or if $y_i = 0$ whenever $d_i = 0$. This practical problem in specific samples particularly arises for dummy variables d_i with a small number of $d_i = 1$ (e.g. if sectoral dummies are included as explanatory variables, but only a small number of firms belongs to a specific industry). These binary explanatory variables have to be dropped which is made by econometric software packages such as STATA by default.
- In the case of linear regression models, heteroskedasticity is not a strong problem since the OLS estimators remain unbiased and consistent and heteroskedasticity robust t statistics can be applied. In contrast, heteroskedasticity in binary response models is a stronger problem. Indeed the ML estimators remain consistent if the heteroskedasticity is unrelated to the explanatory variables. However, if it is related to the explanatory variables, the ML estimators are generally inconsistent so that it could only be accounted for in the log-likelihood function if the form of heteroskedasticity is known.
- As in most econometric models, endogenous variables or endogeneity is also a strong problem in binary response models since the ML estimator is inconsistent in this case. Against this background, several binary response models with endogenous regressors have been developed and discussed.

Binary discrete choice models:

- The analysis of binary response models can be related to microeconomic models of choice by interpreting data on individual choices between the two alternatives j = 0, 1 of the binary dependent variable y_i (e.g. employment or unemployment) within the random utility model
- In the following u_{i0} is the (unobservable) utility of observation i from alternative j = 0 and u_{i1} is the utility from alternative j = 1. It is assumed that i chooses alternative j = 1 if $u_{i1} > u_{i0}$ and alternative j = 0 if $u_{i1} \leq u_{i0}$.
- It is further assumed that the following hypothetical utility function u_{ij} of i for alternative j depends on a vector $x_i = (x_{i1},\ldots, x_{ik_1})'$ of individual characteristics as examined so far, which are constant across both alternatives j = 0, 1 (e.g. age, education), and additionally a vector $z_{ij} = (z_{ij1},\ldots, z_{ijk_2})'$ of alternative specific attributes which can vary across the alternatives (e.g. price for a means of transportation):

$$u_{ij} = \beta_j' x_i + \gamma' z_{ij} + \varepsilon_{ij} \quad \text{for } i = 1,\ldots, n;\ j = 0, 1$$

$\beta_j = (\beta_{j1},\ldots, \beta_{jk_1})'$ and $\gamma = (\gamma_1,\ldots, \gamma_{k_2})'$ are the unknown parameter vectors that have to be estimated and ε_{ij} is an unobservable stochastic component.

- The probabilities for y_i = 0 and y_i = 1 can then be calculated by computing the probabilities that u_{i0} or u_{i1} is higher, respectively. Different binary discrete choice models (such as binary probit and logit models) result from different assumptions for the stochastic component. This discrete choice analysis is particularly important for multinomial response models.