7.3 Models with a qualitative dependent variable

Models with qualitative dependent variables typically appear in an **explanation of economic choices** in dependence on specific characteristics.

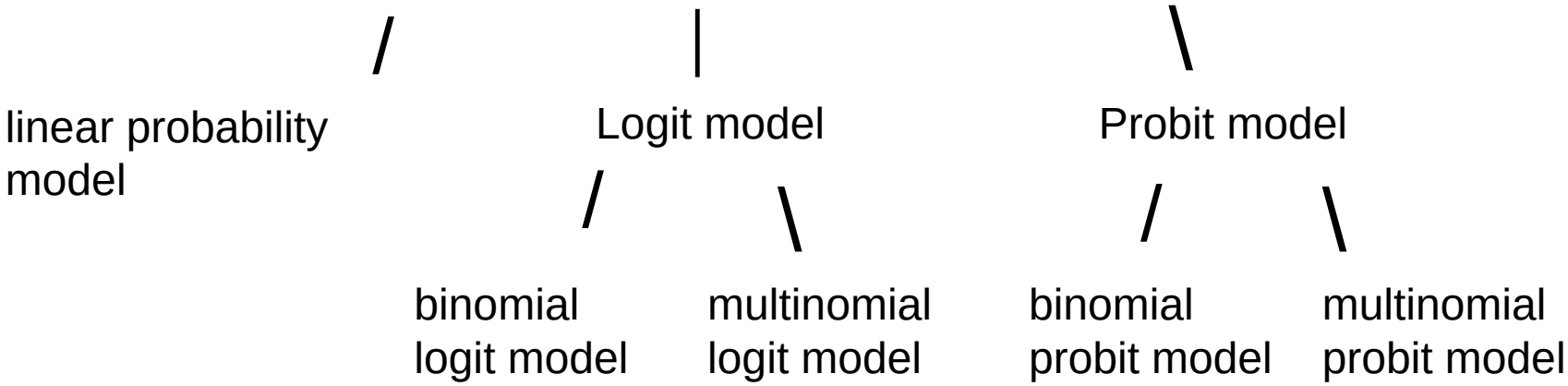
Example:

- With which factors can the decision of an individual to supply his labour be explained economically?
- Which influencing factors are decisive for the decision of a consumer to buy or not to buy a product?
- Whereby is the choice of a means of transport being explained economically?

□

Type of econometric analysis:
Cross-sectional analysis with micro data (microeconometric analysis)

Types of models with qualitative dependent variables



In what follows, we introduce the linear probability model and the logit model. We focus the analysis in particular on the binomial logit model, in which the dependent variable is dichotomous, i.e. it has two attributes such as buying (1) and not buying (0). Multinomial logit models present insofar a generalization in that the dependent variable has more than two attributes such as when choosing a means of transport car (1), train (2), and bus (3).

Logit model: based on the logistic distribution

Probit model: based on the normal distribution

Example:
The purchase of a product in dependence on the income and the gender of a consumer is being studied. From a cross-sectional survey the following database has been generated:

Konsument t	Produktkauf y_t	Einkommen x_{2t}	Geschlecht x_{3t}
1	nein (0)	2500	weibl. (1)
2	ja (1)	3000	männl. (0)
3	ja (1)	3200	männl. (0)
4	ja (1)	3500	männl. (1)
5	nein (0)	2800	weibl. (1)
6	nein (0)	3000	weibl. (1)
7	ja (1)	4000	männl. (0)
8	nein (0)	2200	weibl. (1)
9	ja (1)	3800	männl. (0)
10	nein (0)	2900	weibl. (1)
11	ja (1)	2600	weibl. (1)
12	nein (0)	2400	männl. (0)
13	ja (1)	4200	männl. (0)
14	ja (1)	3700	männl. (0)
15	nein (0)	3300	weibl. (1)
16	ja (1)	4100	weibl. (1)
17	nein (0)	2500	weibl. (1)
18	ja (1)	3900	männl. (0)
19	ja (1)	4400	männl. (0)
20	nein (0)	2000	weibl. (1)
21	nein (0)	3100	weibl. (1)
22	ja (1)	3400	männl. (0)
23	ja (1)	3500	weibl. (1)
24	ja (1)	3800	männl. (0)
25	nein (0)	2200	weibl. (1)

Profile feature of consumer 1, who has not purchased the product (y_1= 0):

$$\mathbf{x}_1' = (1 \quad 2500 \quad 1)$$

Profile feature of consumer 2, who has purchased the product (y_1= 1):

$$\mathbf{x}_2' = (1 \quad 3000 \quad 0)$$

● The linear probability model

The dichotomous dependent variable y assumes the value 1 with the probability p_t if an event such as the purchase of a product is realized or if a person t of working age is employed. If the considered event does not occur, the random variable y takes on the value 0.

While the purchasing decision may depend on quantities such as e.g. income, gender, and social status, are e.g. the wage rate and gender important determinants of a person‘s employment decision.

In general, in a discrete choice model the probability that an individual decides to e.g. a purchase of a product or a participation in working is life explained by the influencing terms x_1, x_2, ..., x_k:

(7.9) $$P(y_t = 1) = p_t = F(\boldsymbol{\beta}'\mathbf{x}_t),\ t = 1,2,\ldots,n$$

with $$\mathbf{x}_t = (x_{1t} \quad x_{2t} \quad \ldots \quad x_{kt})'$$

and $$\boldsymbol{\beta} = (\beta_1 \quad \beta_2 \quad \ldots \quad \beta_k)'$$

The vector β contains the unknown parameters of the choice model, which is accorded a disparate interpretation, depending on the specification of the distribution function F. As a rule is x_1 a pseudo variable (x_{1t}= 1 for all t), so that the parameter β_1 denotes the intercept.

In the simplest choice model, for example, the purchase probability changes linearly with a change in the influencing factors. One obtains the **linear probability model (LP model)** under the assumption that the base distribution is a uniform distribution:

(7.10) $$p_t = F(\boldsymbol{\beta}'\mathbf{x}_t) = \boldsymbol{\beta}'\mathbf{x}_t$$

The probability p_t can be interpreted at the same time as the (conditional) expected value of the random variable y given the vector **x**:

(7.11) $$E(y_t | \mathbf{x}_t) = p_t = \boldsymbol{\beta}'\mathbf{x}_t$$

The random variable y_t is accordingly composed of the systematic component p_t and a disturbance term u_t that takes account of the possible deviations between the observed and expected choice decisions of the individuals:

(7.12) $$y_t = p_t + u_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t$$

Since y_t can only assume the values 1 and 0, u_t follows a two-point distribution, which is made apparant in the following the Table:

y_t	u_t	$P(u_t)$
1	$1-\boldsymbol{\beta}'\mathbf{x}_t$	$\boldsymbol{\beta}'\mathbf{x}_t$
0	$-\boldsymbol{\beta}'\mathbf{x}_t$	$1-\boldsymbol{\beta}'\mathbf{x}_t$

Expected value of u_t:

(7.13) $$E(u_t)=(1-\boldsymbol{\beta}'\mathbf{x}_t)\boldsymbol{\beta}'\mathbf{x}_t-\boldsymbol{\beta}'\mathbf{x}_t(1-\boldsymbol{\beta}'\mathbf{x}_t)=0$$

Variance of u_t:

(7.14) $$\begin{aligned} Var(u_t) &= E(u_t^2)=(1-\beta' x_t)^2\,\beta' x_t+(\beta' x_t)^2(1-\beta' x_t) \\ &= (1-\beta' x_t)\beta' x_t[(1-\beta' x_t)+\beta' x_t] \\ &= (1-\beta' x_t)\beta' x_t \end{aligned}$$

Methodological problems of the LP model:

- No guarantee that the estimated probabilities lie in the interval between 0 and 1,
- The variance of the disturbance term u_t depends on the explanatory variables (x-variables), so that the disturbance term is from the outset heteroskedastic,
- The strength of the impulses on the change in the probability of an event that emanate from the x-variables are independent of the their realized level.

Example:

Ordinary least-squares estimation (OLS estimation) of the linear probability model (LP model):

Dependent Variable: KAUF
Method: Least Squares
Method: Least Squares
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.661990	0.425109	-1.557222	0.1337
EINKOMMEN	0.000434	0.000114	3.795790	0.0010
GESCHLECHT	-0.299123	0.155211	-1.927202	0.0670

R-squared	0.630553	Mean dependent var	0.560000
Adjusted R-squared	0.596967	S.D. dependent var	0.506623
S.E. of regression	0.321629	Akaike info criterion	0.681330
Sum squared resid	2.275794	Schwarz criterion	0.827595
Log likelihood	-5.516625	Hannan-Quinn criter.	0.721898
F-statistic	18.77423	Durbin-Watson stat	2.439003
Prob(F-statistic)	0.000018		

• Observed values of the dependent variables, estimated probabilities, residuals

t	y_t	$\hat{p}_t$	$\hat{u}_t$
1	0.00000	0.12443	-0.12443
2	1.00000	0.64067	0.35933
3	1.00000	0.72751	0.27249
4	1.00000	0.55865	0.44135
5	0.00000	0.25470	-0.25470
6	0.00000	0.34154	-0.34154
7	1.00000	1.07488	-0.07488
8	0.00000	-0.00583	0.00583
9	1.00000	0.98804	0.01196
10	0.00000	0.29812	-0.29812
11	1.00000	0.16786	0.83214
12	0.00000	0.38013	-0.38013
13	1.00000	1.16173	-0.16173
14	1.00000	0.94462	0.05538
15	0.00000	0.47181	-0.47181
16	1.00000	0.81918	0.18082
17	0.00000	0.12443	-0.12443
18	1.00000	1.03146	-0.03146
19	1.00000	1.24857	-0.24857
20	0.00000	-0.09268	0.09268
21	0.00000	0.38496	-0.38496
22	1.00000	0.81435	0.18565
23	1.00000	0.55865	0.44135
24	1.00000	0.98804	0.01196
25	0.00000	-0.00583	0.00583

Estimated probabilities <0:

Consumers 8, 20, 25

Estimated probabilities > 1:

Consumers 7, 13, 18, 19

● The logit model

For these reasons, empirical economic research draws on the **probit and logit models,** instead of the linear probability model, when seeking to explain economic choice actions. While in the probit model, for example, the purchase probability is given by a standard normal distribution, the logit model is based on the distribution function of a logistic distribution. Since the probability functions of the probit and logit models comply in their form, the logit model as the better manageable for estimation purposes I spreferred in empirical studies. The difference is essentially that the logistic probability distribution possesses a greater probability mass at the tails than the normal distribution, which often better corresponds with the empirical evidence.

The **Logit model** is given by the specification of the distribution function F [Eq. (7.9)] in the form of the logistic distribution:

$$(7.15) \quad p_t = F(\boldsymbol{\beta}'\mathbf{x}_t) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_t}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_t}} = \frac{1}{1+e^{-\boldsymbol{\beta}'\mathbf{x}_t}} = \frac{1}{1+e^{-(\beta_1+\beta_2 \cdot x_{2t}+\ldots+\beta_k \cdot x_{kt})}}$$

- Justification for the normal distribution instead of the uniform distribution

The normal distribution can be motivated when one for an individual t introduces the unobservable index variable

$$y_t^* = \boldsymbol{\beta}' \mathbf{x}_t + u_t$$

which captures the degree of predisposition of the individual, for example, for a purchasing decission or work participation. The individual will decide to buy a good if the index variable y_t^* exceeds a certain threshold at which point the possession of the good is considered to be advantageous. If, without loss of generality, we set the threshold equal to zero, the relationship between the unobservable index variable y_t^* and the observable dependent variable y_t can be presented by

$$y_t = 1 \quad \text{falls } y_t^* > 0$$
$$y_t = 0 \quad \text{falls } y_t^* \leq 0$$

The probability that an individual decides, for example. to buy a product or to take up an employment is then given by

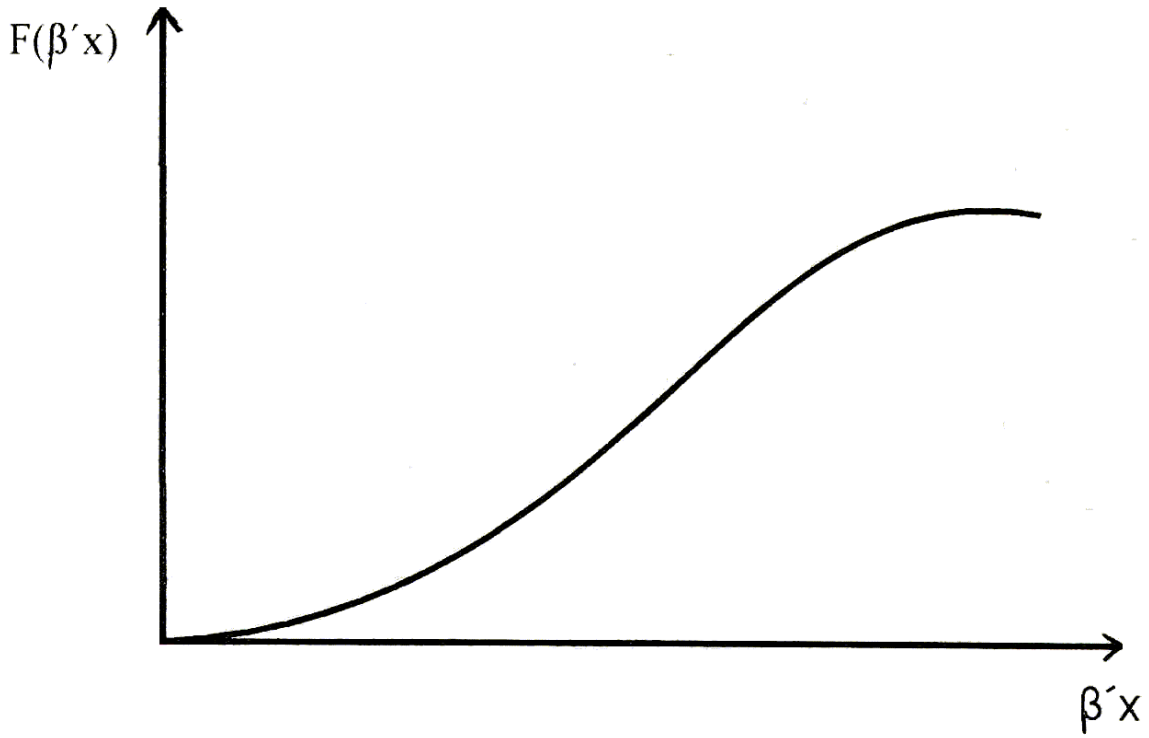
$$P\left(y_t^* > 0\right) = P\left(\boldsymbol{\beta}' \mathbf{x}_t + u_t > 0\right) = P\left(u_t > -\boldsymbol{\beta}' \mathbf{x}_t\right)$$

and with a symmetry of the probability distribution by

$$P\left(y_t^* > 0\right) = P\left(u_t < \boldsymbol{\beta}' \mathbf{x}_t\right) = F\left(\boldsymbol{\beta}' \mathbf{x}_t\right) \ .$$

If one envisions that the index variable y_t^* is made up of a multiplicity of independent influencing factors, and due to the central limit theorem for F the distribution function of a normal distribution can be used. This justifies the utilisation of the **probit model** in the econometric analysis of economic choice actions. If in applications a greater proportion of "outside values" than with a normal distribution can be expected, the **logit model** provides, however, a more suitable modeling basis.

Figure: The probability distribution of the logit model



• Odds and effect coefficients

The odds express the probability ratio

(7.16) $$\frac{P(y_t = 1)}{1 - P(y_t = 1)} = \frac{p_t}{1 - p_t}$$

i.e. the odds indicate the ratio of the probability of a realisation of the event $y_t = 1$ relative to the probability of a realization of the contrary event $y_t = 0$.

In the logit analysis, the odds are given by

(7.17) $$\frac{p_t}{1 - p_t} = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_t}}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_t}} \Bigg/ \frac{1}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_t}} = e^{\boldsymbol{\beta}'\mathbf{x}_t}$$

Thus the effect coefficient e^{β_j} indicates how the odds change if the explanatory variable x_j is increased by one unit:

$$e^{\beta_1 + \beta_2 \cdot x_{2t} + \ldots + \beta_j \cdot (x_{jt} + 1) + \ldots + \beta_j \cdot x_{kt}}$$

$$= \underbrace{e^{\beta_1 + \beta_2 \cdot x_{2t} + \ldots + \beta_j \cdot x_{jt} + \ldots + \beta_j \cdot x_{kt}}}_{\text{Odds}} \cdot e^{\beta_j} = \frac{p_t}{1 - p_t} \cdot e^{\beta_j}$$

After taking the logarithm of (7.17) one gets the logit model in the form

(7.18) $$\ln \frac{p_t}{1-p_t} = \boldsymbol{\beta}' \mathbf{x}_t$$

from which can be seen that the regression coefficients β_j of the logit model measure the impulses on the so-called log odds $\ln(p_t/1-p_t)$, which result from a change in the exogenous factors by one unit. The log odds display the chance of a realization of the considered event in comparison to the complementary event. For this, the denotation "logit" has been adopted into the language.

• Marginal effect of the explanatory variables x_j on the dependent variable y

The influences of a change in the explanatory variables on the probability of the observed event can be specified by the partial derivatives

(7.19) $$\frac{\partial p_t}{\partial x_{jt}} = \frac{\partial F(\boldsymbol{\beta}' \mathbf{x}_t)}{\partial x_{jt}} = f(\boldsymbol{\beta}' \mathbf{x}_t) \cdot \beta_j$$

, in which f denotes the density function of the logit distribution:

$$f(\boldsymbol{\beta}' \mathbf{x}_t) = \frac{e^{\boldsymbol{\beta}' \mathbf{x}_t}}{\left(1 + e^{\boldsymbol{\beta}' \mathbf{x}_t}\right)^2}$$

The coefficients β_j only indicage the direction in which the probabilities p_t change by certain impulses of the exogenous variables. The magnitude of the change, however, depends also on the steepness of the distribution function of the logit distribution at the point $\boldsymbol{\beta}' \mathbf{x}_t$. The steeper the distribution function at the prevailing level is, the greater is the effect of a change in the explanatory variables on the probability of the considered event.

- Maximum likelihood estimation of the logit model

p_t: Probability that the random variable y_t takes the value 1

with validity of the logit model: $p_t = e^{\boldsymbol{\beta}'\mathbf{x}_t} \Big/ 1 + e^{\boldsymbol{\beta}'\mathbf{x}_t}$ (Eq. 7.15)

Vector $\mathbf{x}_t$ of the explanatory variables: characteristic profiles of the individuals

→ Estimation of the logit model with the **maximum likelihood method**

Principle of maximum likelihood (ML) method:

From what population originates presumably a concrete sample? If one varies the parameters of the population, the probability that the present sample is realized will also change. Under which parameter values will the given sample achieve the maximum probability to become realized? These values are called **maximum likelihood estimator (ML-estimator)**.

Sample of the values (0 or 1) of the dependent variable y: $y_1, y_2, ..., y_n$

Probability that for individual t y_t=1 or y_t=0 occurs:

$$f(y_t) = p_t^{y_t} \cdot (1 - p_t)^{1-y_t}$$

for y_t= 1: f (y_t= 1) = p_t, for y_t= 0: f (y_t= 0) = 1-p_t

Joint probability function f_y for $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_n = y_n$ obeying a Bernoulli-process (independent draws):

(7.16)
$$f_y(y_1, y_2, ..., y_n | \boldsymbol{\beta}, \mathbf{x})$$
$$= p_1^{y_1} \cdot (1-p_1)^{1-y_1} \cdot p_2^{y_2} \cdot (1-p_2)^{1-y_2} \cdot ... \cdot p_n^{y_n} \cdot (1-p_n)^{1-y_n}$$

Likelihood function:

(7.17)
$$L(\boldsymbol{\beta} \mid \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n, \mathbf{y}) = \prod_{t=1}^{n} p_t^{y_t} (1-p_t)^{1-y_t} = \prod_{t=1}^{n} \left(\frac{e^{\boldsymbol{\beta}' \mathbf{x}_t}}{1+e^{\boldsymbol{\beta}' \mathbf{x}_t}} \right)^{y_t} \left(1 - \frac{e^{\boldsymbol{\beta}' \mathbf{x}_t}}{1+e^{\boldsymbol{\beta}' \mathbf{x}_t}} \right)^{1-y_t}$$

For the purpose of estimating the parameter vector β we consider the logarithmic likelihood function L*:

(7.18)
$$L^* = \ln L = \sum_{t=1}^{n} y_t \cdot \ln p_t + (1-y_t) \cdot \ln(1-p_t)$$
$$= \sum_{t=1}^{n} y_t \cdot \ln \frac{e^{\beta' x_t}}{1+e^{\beta' x_t}} + (1-y_t) \cdot \ln \frac{1}{1+e^{\beta' x_t}}$$

First-order condition for a maximum of the likelihood function:

$$\frac{\partial L^*}{\partial \boldsymbol{\beta}} = \sum_{t=1}^{n} \frac{y_t}{p_t}\left(\frac{e^{\boldsymbol{\beta}'\mathbf{x}_t}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_t}}\right)\mathbf{x}_t + (1-y_t)\frac{1}{1-p_t}\frac{-e^{\boldsymbol{\beta}'\mathbf{x}_t}}{\left(1+e^{\boldsymbol{\beta}'\mathbf{x}_t}\right)^2}\mathbf{x}_t = 0$$

$$= \sum_{t=1}^{n}(y_t - p_t)\,\mathbf{x}_t = \mathbf{0}$$

(7.19) $$= \sum_{t=1}^{n} \cdot [y_t - p_t(\boldsymbol{\beta})] \cdot \mathbf{x}_t = \mathbf{0} \quad \text{with } p_t(\beta) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_t}}{1+e^{\boldsymbol{\beta}'\mathbf{x}_t}}$$

Since (7.19) is a non-linear system of equations in **β**, the unknown parameter vector cannot be analytically determined. Rather an iterative approach is to be used with which the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{ML}$ for **β** can be determined numerically.

Such an iteration procedure, possessing a high speed of convergence, is the Newton-Raphson procedure.

Starting from an initial estimator $\boldsymbol{\beta}^{(0)}$ with the Newton-Raphson procedure in the m-th step an improved parameter estimator is being identified from an iterative scheme of the form :

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left[\frac{\partial^2 L^*}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'} \Big| \boldsymbol{\beta} = \boldsymbol{\beta}^{(m)} \right)^{-1} \left(\frac{\partial L^*}{\partial \boldsymbol{\beta}} \Big| \boldsymbol{\beta} = \boldsymbol{\beta}^{(m)} \right) \Bigg]$$

The iteration specification is based on a second-order Taylor series decomposition. For the logit model the vector of the first derivatives (= gradient) and the matrix of the second derivatives (= Hessian matrix) are given by

$$\frac{\partial L^*}{\partial \boldsymbol{\beta}} = \sum_{t=1}^{n} \left[y_t \frac{1}{1+e^{\boldsymbol{\beta}' \mathbf{x}_t}} - (1-y_t) \frac{1}{1+e^{\boldsymbol{\beta}' \mathbf{x}_t}} \right] \cdot \mathbf{x}_t$$

and

$$\frac{\partial^2 L^*}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'} = \sum_{t=1}^{n} \frac{e^{-\boldsymbol{\beta}' \mathbf{x}_t}}{\left(1+e^{-\boldsymbol{\beta}' \mathbf{x}_t}\right)^2} \cdot \mathbf{x}_t \mathbf{x}_t'$$

.

At the same time the Hessian inverse can be used as an estimator for the covariance matrix of the ML–estimator $\hat{\boldsymbol{\beta}}_{ML}$. Its diagonal elements, which contain the variances of the components of $\hat{\boldsymbol{\beta}}_{ML}$, can be utilized for a significance test of the estimated coefficients of the logit model.

Example:

Logit model for the product purchase (y) depending on income (x_2) and gender (x_3) reads:

$$p_t = \frac{e^{\beta_1+\beta_2\cdot x_{2t}+\beta_3\cdot x_{3t}}}{1+e^{\beta_1+\beta_2\cdot x_{2t}+\beta_3\cdot x_{3t}}} = \frac{1}{1+e^{-(\beta_1+\beta_2\cdot x_{2t}+\beta_3\cdot x_{3t})}}$$

- Maximum likelihood estimation (ML-estimation) of the logit model

Part 1: Regression coefficients with standard errors, z and p values

Dependent Variable: KAUF
Method: ML - Binary Logit (Quadratic hill climbing)
Sample: 1 25 Included observations: 25
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-11.72644	5.960073	-1.967499	0.0491
EINKOMMEN	0.004480	0.002038	2.197920	0.0280
GESCHLECHT	-2.874458	1.836442	-1.565232	0.1175

- Estimated logit model for the product purchase

$$\hat{p}_t = \frac{1}{1+e^{-(\hat{\beta}_1+\hat{\beta}_2\cdot x_{2t}+\hat{\beta}_3\cdot x_{3t})}} = \frac{1}{1+e^{11{,}72644-0{,}00448x_{2t}+2{,}87446x_{3t}}}$$

Effects of the explanatory variables income (x_2) and gender (x_3) on the purchase probability p:

$\hat{\beta}_2 = 0{,}00448$ > 0: Income (x_2) ↑ → Purchase probability p ↑

$\hat{\beta}_3 = -2{,}87446$ <0: gender (x_3) female (1) instead of male (0)→ purchase prob. p↑

Probability of purchase for consumer 1, who has decided for the non-purchase (y_1= 0), by virtue of her characteristics profile $\mathbf{x}'_1$ = (1 2500 1):

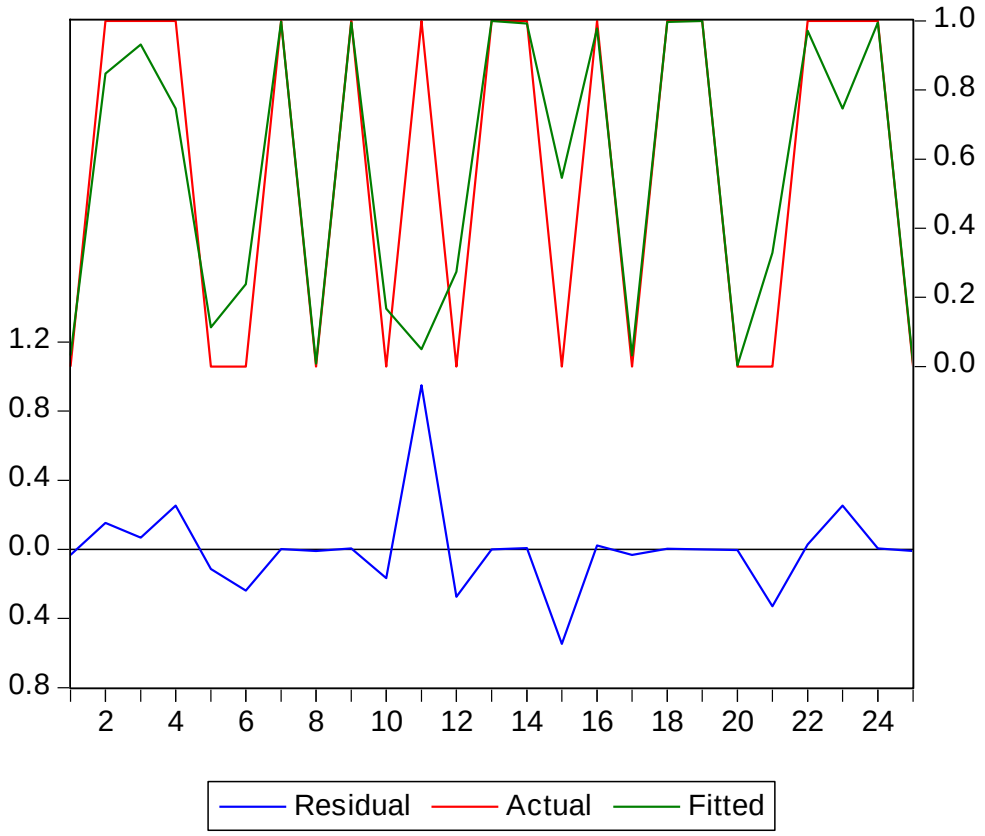
$$\hat{p}_1 = \frac{1}{1+e^{11{,}72644-0{,}00448\cdot 2500+2{,}87446\cdot 1}} = 0{,}0323$$

Probability of purchase for consumer 2, who has decided for the purchase (y_2=1), by virtue of his characteristics profile $\mathbf{x}'_2$ = (1 3000 0):

$$\hat{p}_2 = \frac{1}{1+e^{11{,}72644-0{,}00448\cdot 3000+2{,}87446\cdot 0}} = 0{,}8473$$

- Observed values of the dependent variables, estimated probabilities, residuals

t	y_t	$\hat{p}_t$	$\hat{u}_t$
1	0.00000	0.03230	-0.03230
2	1.00000	0.84746	0.15254
3	1.00000	0.93156	0.06844
4	1.00000	0.74659	0.25341
5	0.00000	0.11347	-0.11347
6	0.00000	0.23873	-0.23873
7	1.00000	0.99796	0.00204
8	0.00000	0.00863	-0.00863
9	1.00000	0.99503	0.00497
10	0.00000	0.16691	-0.16691
11	1.00000	0.04965	0.95035
12	0.00000	0.27419	-0.27419
13	1.00000	0.99917	0.00083
14	1.00000	0.99224	0.00776
15	0.00000	0.54598	-0.54598
16	1.00000	0.97744	0.02256
17	0.00000	0.03230	-0.03230
18	1.00000	0.99682	0.00318
19	1.00000	0.99966	0.00034
20	0.00000	0.00354	-0.00354
21	0.00000	0.32924	-0.32924
22	1.00000	0.97088	0.02912
23	1.00000	0.74659	0.25341
24	1.00000	0.99503	0.00497
25	0.00000	0.00863	-0.00863



• Variance-covariance-matrix of the regression coefficients

- calculated as the inverse of the matrix of the second derivatives (= Hessian inverse)

$$\left.\frac{\partial^2 L^*}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{ML}} = \sum_{t=1}^{n} \frac{e^{-\hat{\boldsymbol{\beta}}'\mathbf{x}_t}}{\left(1+e^{-\hat{\boldsymbol{\beta}}'\mathbf{x}_t}\right)^2} \cdot \mathbf{x}_t \mathbf{x}_t'$$

(Hessian matrix at the location $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{ML}$)

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{ML}) = \left(\left.\frac{\partial^2 L^*}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{ML}} \right)^{-1} = \begin{bmatrix} 35.52247 & -0.011758 & 1.477915 \\ -0.011758 & 4.16E\text{-}06 & -0.001318 \\ 1.477915 & -0.001318 & 3.372518 \end{bmatrix}$$

Standard error of the regression coefficients:

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{35{,}52247} = 5{,}960073$$

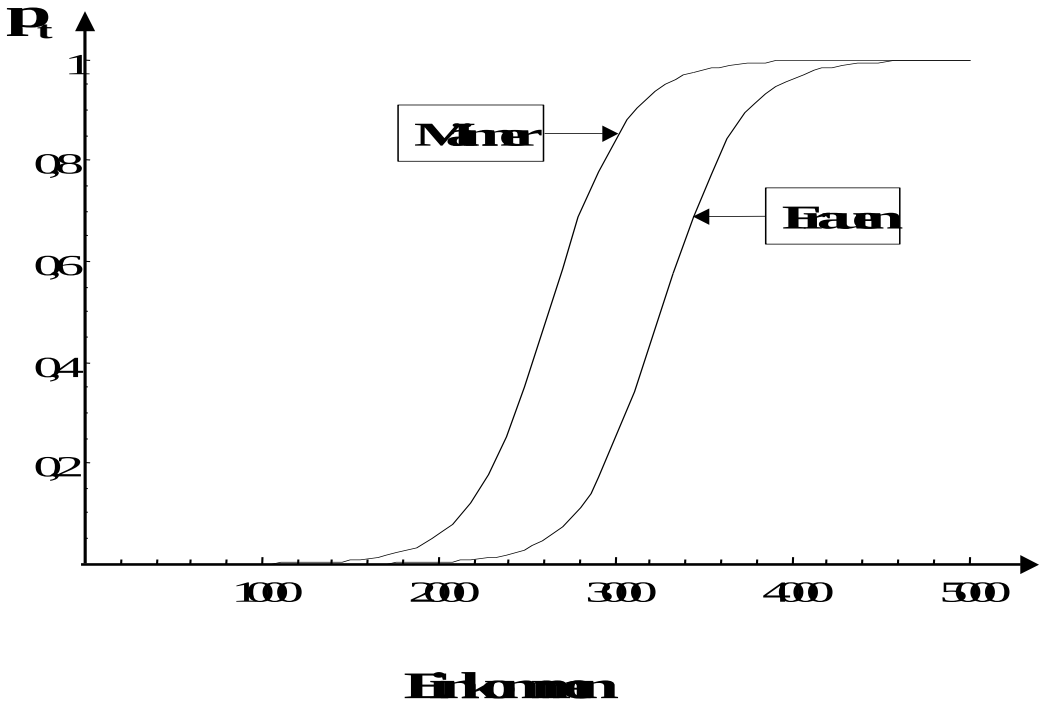
$$\hat{\sigma}_{\hat{\beta}_2} = \sqrt{0{,}00000416} = 0{,}0020240$$

$$\hat{\sigma}_{\hat{\beta}_3} = \sqrt{3{,}372518} = 1{,}836442$$

• Behaviour of the estimated cumulative purchasing probabilities for women and men in dependence on income (x_2)

for women: $$\hat{p}_t = \frac{1}{1 + e^{11{,}72644 - 0{,}00448 \cdot x_{2t} + 2{,}87446}}$$

for men: $$\hat{p}_t = \frac{1}{1 + e^{11{,}72644 - 0{,}00448 \cdot x_{2t}}}$$



From the graph it also becomes apparent that the regression coefficients can not be interpreted in just the same way as in the linear regression. If income in the middle range at 3,000 € increases by one unit, then the purchasing probability rises much steeper than with an equally strong change in the edge regions (for example, at 1,000 € or at 5,000 €).

- Marginal effect of income (x_2) on the purchasing probability p

$$\frac{\partial \hat{p}_t}{\partial x_{2t}} = f\left(\hat{\boldsymbol{\beta}}' \mathbf{x}_t\right) \cdot \hat{\beta}_2 = \frac{e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_t}}{\left(1 + e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_t}\right)^2} \cdot \hat{\beta}_2$$

- Effect of an increase of income by 100 € at consumer 1 (t=1) on the probability of purchasing the product:

$$\frac{\partial \hat{p}_t}{\partial x_{21}} \cdot 100 = \frac{e^{-11,72644+0,00448 \cdot 2500-2,87446 \cdot 1}}{\left(1 + e^{-11,72644+0,00448 \cdot 2500-2,87446 \cdot 1}\right)^2} \cdot 0,00448 \cdot 100$$

$$= 0,03123 \cdot 0,00448 \cdot 100 = 0,0140$$

- Effect of an increase in income by 100 € at consumer 2 (t=2) on the probability of purchasing the product:

$$\frac{\partial \hat{p}_t}{\partial x_{22}} \cdot 100 = \frac{e^{-11,72644+0,00448 \cdot 3000-2,87446 \cdot 0}}{\left(1 + e^{-11,72644+0,00448 \cdot 3000-2,87446 \cdot 0}\right)^2} \cdot 0,00448 \cdot 100$$

$$= 0,12938 \cdot 0,00448 \cdot 100 = 0,0580$$

• Effect coefficients (odds ratios)

- Estimated effect coefficient of income (x_2)

$e^{\hat{\beta}_2} = e^{0,00448} = 1,0045 \quad > 1 \rightarrow$

With increasing incomes the odds rise, i.e. the chance ratio for the product purchase compared to the non-purchase increases.

$(e^{0,00448} - 1) \cdot 100\% = 0,45\% \quad \rightarrow$

If income increases by 1 GE, then the odds within the group of female and male consumers increase by a factor of 1,0045, i.e. by 0,45%.

- Estimated effect coefficient of gender (x_3)

$e^{\hat{\beta}_3} = e^{-2,87446} = 0,0564 \quad < 1 \rightarrow$

With a transition from the group of male consumers (x_3=0) to the group of female consumers (x_3=1) the odds decline, i.e. the chance ratio for the product purchase compared to the non-purchase declines.

$(e^{-2,87446} - 1) \cdot 100\% = 94,36\% \quad \rightarrow$

If one transits from the group of male consumers (x_2=0) to the group of female consumers (x_3 = 1) while keeping income constant, the odds decline by a factor of 17,7, i.e. by 94.36%.

● Likelihood ratio test and pseudo-R^2

The general validity of the logit model can be appraised using the **likelihood ratio test**. If the null hypothesis

(7.20) $H_0: \quad \beta_2 = \beta_3 = \ldots = \beta_p = 0$

is valid, the probability that the random variable y assumes the value 1 can be estimated, in that one relates the number of statistical units n_1, that have the considered characteristic, to the number of observations n:

$$\hat{p} = \frac{n_1}{n} \rightarrow p(y = 1) = p$$

A conditionality by exogenous variables is dispensed, because it does not possess any explanatory content. Under the null hypothesis the logarithmic likelihood function is given by

(7.21) $$L_0^* = \ln L_0 = n_1 \ln\left(\frac{n_1}{n}\right) + (n - n_1)\ln\left(\frac{n - n_1}{n}\right)$$

While L_0^* is denoted the restricted log-likelihood function, the logarithmic likelihood function L* with the logit model is called unrestricted. Under the null hypothesis (7.20) the likelihood ratio statistic

(7.22) $$l = -2\left(L_0^* - L^*\right)$$

follows asymptotically a χ^2-distribution with k-1 degrees of freedom. An acceptance of the null hypothesis implies that exogenous variables x_2, x_3, ..., x_k have no influence on the probability of the observed event, so that their estimation is solely justified by the sample faction $\hat{p}$.

As a measure of the goodness of fit a measure comparable to the coefficient of determination has been developed that is called **pseudo-R^2**:

(7.23) $$PR^2 = 1 - \frac{L^*}{L_0^*}$$

The pseudo-R^2 can be understood as a measure of quality that gauges the degree of "uncertainty" in the data which is "explained" by the logit model. If p_t is always 1 if y_t= 1 and 0 when y_t= 0, then for the unrestricted log–likelihood function L*=ln1=0 applies, i.e. PR^2 then reaches its maximum value. In this extreme case the explanatory variables provide a perfect model fit. Obviously this situation can only occur, however, if $\boldsymbol{\beta}'\mathbf{x}_t$ approaches -∞ and +∞. The minimum value of the pseudo-R^2 is achieved exactly when the unrestricted log–likelihood function coincides with the restricted, i.e. when the exogenous variables have no explanatory power whatsoever. Even though it is basically possible with growing PR^2 to speak of an improved model fit, still the values of the pseudo-R^2 do not have the vivid interpretation, which holds for the coefficient of determination in the multiple regression model.

Example:

Parameters and tests on the model fit for the estimated logit model:

McFadden R-squared	0.645914	Mean dependent var	0.560000
S.D. dependent var	0.506623	S.E. of regression	0.273253
Akaike info criterion	0.725756	Sum squared resid	1.642677
Schwarz criterion	0.872021	Log likelihood	-6.071946
Hannan-Quinn criter.	0.766323	Restr. log likelihood	-17.14825
LR statistic	22.15260	Avg. log likelihood	-0.242878
Prob(LR statistic)	0.000015		

Obs with Dep=0	11	Total obs	25
Obs with Dep=1	14		

- Pseudo - R^2

  - Value of the log likelihood function under the null hypothesis (restricted log likelihood function L_0^*):

$$L_0^* = \ln L_0 = n_1 \ln\left(\frac{n_1}{n}\right) + (n - n_1)\ln\left(\frac{n - n_1}{n}\right)$$

$$= 14 \cdot \ln\left(\frac{14}{25}\right) + 11 \cdot \ln\left(\frac{11}{25}\right) = -17{,}148$$ (Eviews output: Restr. log likelihood.)

  - Value of the log likelihood function for the estimated logit model (un-restricted log likelihood function L*):

$$L^* = -6{,}072$$ (EViews output: Log likelihood)

(Calculation see next page)

$$PR^2 = 1 - \frac{L^*}{L_0^*} = 1 - \frac{-6{,}072}{-17{,}148} = 0{,}646$$ (EViews Output: McFadden R-squared)

Calculation of the unrestricted log likelihood function L*:

t	y_t	$p_t = \hat{y}_t$	$y_t \cdot \ln p_t$	$(1-y_t)\cdot \ln(1-p_t)$	$y_t \cdot \ln p_t + (1-y_t)\cdot \ln(1-p_t)$
1	0	0,032300	0,0000	-0,0328	-0,0328
2	1	0,847456	-0,1655	0,0000	-0,1655
3	1	0,931558	-0,0709	0,0000	-0,0709
6	1	0,746596	-0,2922	0,0000	-0,2922
4	0	0,113473	0,0000	-0,1204	-0,1204
5	0	0,238728	0,0000	-0,2728	-0,2728
7	1	0,997965	-0,0020	0,0000	-0,0020
8	0	0,008629	0,0000	-0,0087	-0,0087
9	1	0,995029	-0,0050	0,0000	-0,0050
10	0	0,166908	0,0000	-0,1826	-0,1826
13	1	0,049650	-3,0027	0,0000	-3,0027
11	0	0,274193	0,0000	-0,3205	-0,3205
12	1	0,999168	-0,0008	0,0000	-0,0008
14	1	0,992241	-0,0078	0,0000	-0,0078
15	0	0,545984	0,0000	-0,7896	-0,7896
16	1	0,977440	-0,0228	0,0000	-0,0228
17	0	0,032300	0,0000	-0,0328	-0,0328
18	1	0,996818	-0,0032	0,0000	-0,0032
19	1	0,999660	-0,0003	0,0000	-0,0003
20	0	0,003540	0,0000	-0,0035	-0,0035
21	0	0,329240	0,0000	-0,3993	-0,3993
22	1	0,970885	-0,0295	0,0000	-0,0295
23	1	0,746596	-0,2922	0,0000	-0,2922
24	1	0,995029	-0,0050	0,0000	-0,0050
25	0	0,008629	0,0000	-0,0087	-0,0087
Σ	14	-	-3,9001	-2,1718	$L^* = -6,0719$

- Likelihood ratio test

Null hypothesis H_0: $\beta_2 = \beta_3 = 0$

Test statistic (likelihood ratio statistic):

$$l = -2 \cdot \left(L_0^* - L^* \right) = -2 \cdot \left[-17{,}148 - \left(-6{,}072 \right) \right] = 22{,}152$$

Critical value ($\alpha = 0{,}05$):

$X^2_{k-1,\,1-\alpha} = X^2_{2\;0{,}95} = 5{,}991$

Test decision:

$(L = 22{,}152) > (X^2_{2\;0{,}95} = 5{,}991)$ => reject H_0

- Classification tables

Dependent Variable: KAUF
Method: ML - Binary Logit (Quadratic hill climbing)
Sample: 1 25 Included observations: 25
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	10	1	11	0	0	0
P(Dep=1)>C	1	13	14	11	14	25
Total	11	14	25	11	14	25
Correct	10	13	23	0	14	14
% Correct	90.91	92.86	92.00	0.00	100.00	56.00
% Incorrect	9.09	7.14	8.00	100.00	0.00	44.00
Total Gain*	90.91	-7.14	36.00			
Percent Gain**	90.91	NA	81.82			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	9.25	1.75	11.00	4.84	6.16	11.00
E(# of Dep=1)	1.75	12.25	14.00	6.16	7.84	14.00
Total	11.00	14.00	25.00	11.00	14.00	25.00
Correct	9.25	12.25	21.49	4.84	7.84	12.68
% Correct	84.06	87.47	85.97	44.00	56.00	50.72
% Incorrect	15.94	12.53	14.03	56.00	44.00	49.28
Total Gain*	40.06	31.47	35.25			
Percent Gain**	71.53	71.53	71.53			

*Change in "% Correct" from default (constant probability) specification
**Percent of incorrect (default) prediction corrected by equation