## 7.3 Modell mit qualitativer abhängiger Variablen

Modelle mit qualitativen abhängigen Variablen ergeben sich typischerweise bei einer **Erklärung von ökonomischen Wahlhandlungen** in Abhängigkeit von bestimmten Charakteristika.

**Beispiel:**

- Durch welche Faktoren lässt sich die Entscheidung eines Individuums, seine Arbeitskraft anzubieten, ökonomisch erklären?
- Welche Einflussgrößen sind maßgeblich für den Entschluss eines Konsumenten für den Kauf oder Nichtkauf eines Produkts?
- Wodurch wird die Wahl eines Verkehrs- oder Transportmittels ökonomisch erklärt?

□

Art der ökonometrischen Analyse:
Querschnittsanalyse mit Mikrodaten (mikroökonometrische Analyse)

Arten von Modellen mit qualitativer abhängiger Variablen

- Lineares Wahrscheinlichkeitsmodell
- Logit-Modell
  - Binomiales Logit-Modell
  - Multinomiales Logit-Modell
- Probit-Modell
  - Binomiales Probit-Modell
  - Multinomiales Probit-Modell

Im Folgenden stellen wir das lineare Wahrscheinlichkeitsmodell und das Logit-Modell vor. Speziell fokussieren wir die Analyse auf das binomiale Logit-Modell, bei dem die abhängige Variable dichotom ist, d.h. zwei Ausprägungen wie z.B. Kauf (1) und Nichtkauf (0) hat. Multinomiale Logit-Modelle stellen insofern eine Verallgemeinerung dar, als dass die abhängige Variable mehr als zwei Ausprägungen hat wie z.B. bei der Wahl eines Verkehrsmittels Pkw (1), Bahn (2) und Bus (3).

Logit-Modell: basiert auf der logistischen Verteilung

Probit-Modell: basiert auf der Normalverteilung

**Beispiel:**

Der Kauf eines Produkts wird in Abhängigkeit vom Einkommen und Geschlecht eines Konsumenten untersucht. Aus einer Querschnittserhebung hat sich folgende Datenbasis ergeben:

| 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) |

Merkmalsprofil des Konsumenten 1, der das Produkt nicht gekauft hat ($y_1$=0):

$$\mathbf{x}_1' = \begin{pmatrix} 1 & 2500 & 1 \end{pmatrix}$$

Merkmalsprofil des Konsumenten 2, der das Produkt gekauft hat ($y_1$=1):

$$\mathbf{x}_2' = \begin{pmatrix} 1 & 3000 & 0 \end{pmatrix}$$

● **Lineares Wahrscheinlichkeitsmodell**

Die dichotome abhängige Variable y nimmt den Wert 1 mit der Wahrscheinlichkeit $p_t$ an, falls ein Ereignis wie z.B. der Kauf eines Produkts realisiert wird oder die Erwerbsperson t erwerbstätig ist. Falls das betrachtete Ereignis nicht eintritt, nimmt die Zufallsvariable y den Wert 0 an.

Während die Kaufentscheidung von Größen wie z.B. dem Einkommen, dem Geschlecht und der sozialen Stellung abhängen mag, sind z.B. der Lohnsatz und das Geschlecht maßgebliche Determinanten für die Entscheidung über die Erwerbstätigkeit einer Person.

Allgemein wird bei einem Wahlhandlungsmodell die Wahrscheinlichkeit dafür, dass sich ein Individuum z.B. für den Kauf eines Produkts oder eine Partizipation am Erwerbsleben entscheidet, durch die Einflussgrößen $x_1$, $x_2$, ..., $x_k$ erklärt:

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

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

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

Der Vektor $\boldsymbol{\beta}$ enthält die unbekannten Parameter des Wahlhandlungsmodells, denen in Abhängigkeit von der Spezifikation der Verteilungsfunktion F eine unterschiedliche Interpretation zukommt. In der Regel ist $x_1$ eine Scheinvariable ($x_{1t}$=1 für alle t), so dass der Parameter $\beta_1$ das absolute Glied bezeichnet.

Im einfachsten Wahlhandlungsmodell verändert sich z.B. die Kaufwahrscheinlichkeit linear mit einer Veränderung der Einflussgrößen. Man erhält das **lineare Wahrscheinlichkeitsmodell (linear probability (LP) model)**, in dem als Basisverteilung eine Gleichverteilung unterstellt wird:

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

Die Wahrscheinlichkeit $p_t$ ist bei gegebenem Vektor **x** gleichzeitig der (bedingte) Erwartungswert der Zufallsvariablen y interpretieren:

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

Die Zufallsvariable $y_t$ setzt sich demnach aus der systematischen Komponente $p_t$ und einer Störgröße $u_t$ zusammen, die den möglichen Abweichungen zwischen den beobachteten und erwarteten Wahlhandlungsentscheidungen der Individuen Rechnung trägt:

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

Da $y_t$ nur die Werte 1 und 0 annehmen kann, folgt $u_t$ einer Zweipunktverteilung, die aus der folgenden Tabelle ersichtlich ist:

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

Erwartungswert von $u_t$:

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

Varianz von $u_t$:

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

Methodische Probleme des LP Modells:

- Keine Gewähr, dass die geschätzten Wahrscheinlichkeiten im Intervall zwischen 0 und 1 liegen,
- Die Varianz der Störgröße $u_t$ hängt von den erklärenden Variablen (x-Variablen ) ab, so dass die Störgröße von vornherein heteroskedastisch ist,
- Die Stärke der Impulse, die von den x-Variablen auf die Veränderung der Wahrscheinlichkeit eines Ereignisses ausgehen, sind unabhängig vom ihrem realisierten Niveau.

**Beispiel:**

Gewöhnliche Kleinst-Quadrate-Schätzung (OLS-Schätzung) des linearen Wahrscheinlichkeitsmodells (LP-Modells):

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 | | |

- Beobachtungswerte der abh. Variablen, geschätzte Wahrscheinl., Residuen

| $t$ | $y_t$ | $\hat{p}_t$ | $\hat{u}_t$ |
|---|---|---|---|
| 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 |

Geschätzte Wahrscheinlichkeiten < 0: .....

Konsumenten 8, 20, 25

Geschätzte Wahrscheinlichkeiten > 1: .....

Konsumenten 7, 13, 18, 19

## ● Logit-Modell

In der empirischen Wirtschaftsforschung werden aus diesen Gründen anstelle des linearen Wahrscheinlichkeitsmodells **Probit- und Logit-Modell** zur Erklärung ökonomischer Wahlhandlungen herangezogen. Während im Probit-Modell z.B. die Kaufwahrscheinlichkeit durch eine Standard-Normalverteilung gegeben ist, liegt dem Logit-Modell die Verteilungsfunktion einer logistischen Verteilung zugrunde. Da sich die Wahrscheinlichkeitsfunktionen des Probit- und Logit-Modells in ihrer Form entsprechen, wird das schätztechnisch besser handhabbare Logit-Modell bei empirischen Untersuchungen bevorzugt. Der Unterschied besteht im Wesentlichen darin, dass bei die logistische Wahrscheinlichkeitsverteilung an den Rändern eine größere Wahrscheinlichkeitsmasse als die Normalverteilung besitzt, was der Empirie oft besser entspricht.

Das **Logit-Modell** ist durch die Spezifikation der Verteilungsfunktion F [Gl. (7.9)] in Form der logistischen Verteilung gegeben:

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

- Begründung für die Normalverteilung anstelle der Gleichverteilung

Die Normalverteilung lässt sich begründen, wenn man für ein Individuum t die unbeobachtbare Indexvariable

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

einführt, die den Grad der Neigung des Individuums z.B. für eine Kaufentscheidung oder Arbeitspartizipation erfasst. Das Individuum wird sich für den Kauf eines Gutes entscheiden, wenn die Indexvariable $y_t^*$ eine bestimmte Schranke überschreitet, bei der der Besitz des Gutes als vorteilhaft angesehen wird. Wenn wir die Schranke ohne Einschränkung der Allgemeinheit gleich Null setzen, dann lässt sich die Beziehung zwischen den unbeobachtbaren Indexvariablen $y_t^*$ und der beobachtbaren abhängigen Variablen $y_t$ durch

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

wiedergeben. Die Wahrscheinlichkeit, dass sich ein Individuum z.B. für den Kauf eines Produkts oder die Aufnahme einer Erwerbstätigkeit entscheidet, ist dann durch

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

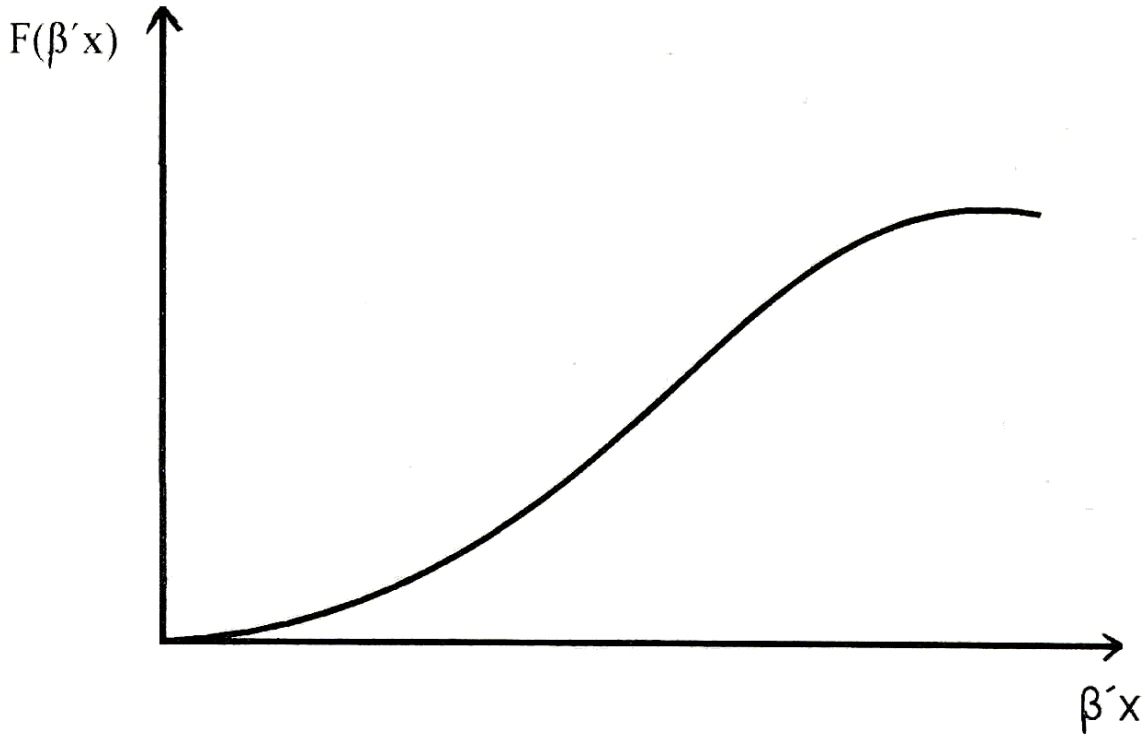
und bei einer Symmetrie der Wahrscheinlichkeitsverteilung durch

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

gegeben.

Wenn man sich vorstellt, dass sich die Indexvariable $y_t^*$ aus einer Vielzahl unabhängiger Einflussfaktoren zusammensetzt, kann aufgrund des Zentralen Grenzwertsatzes für F die Verteilungsfunktion einer Normalverteilung zugrunde gelegt werden. Dies begründet die Verwendung des **Probit-Modells** in der ökonometrischen Analyse von ökonomischen Wahlhandlungen. Wenn bei Anwendungen ein größerer Anteil von "outside values" als bei einer Normalverteilung erwartet werden kann, bietet das **Logit-Modell** jedoch eine geeignetere Modellierungsbasis.

Abbildung: Wahrscheinlichkeitsverteilung des Logit-Modells



• Odds und Effektkoeffizienten

Die Odds drücken das Wahrscheinlichkeitsverhältnis

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

aus, d.h. die Odds geben das Verhältnis der Wahrscheinlichkeit für eine Realisierung des Ereignisses $y_t$=1 relativ zur Wahrscheinlichkeit für eine Realisierung des Gegenereignisses $y_t$=0 wieder.

In der Logit-Analyse sind die Odds durch

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

gegeben. Damit geben die Effektkoeffizienten $e^{\beta_j}$ an, wie sich die Odds verändern, wenn die erklärende Variable $x_j$ um eine Einheit erhöht wird:

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

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

Nach Logarithmierung von (7.17) erhält man das Logit-Modell in der Form

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

aus der ersichtlich ist, dass die Regressionskoeffizienten $\beta_j$ beim Logit-Modell die Impulse auf die sog. Log Odds $\ln(p_t/1-p_t)$ messen, die aus einer Verän-derung der exogenen Grö-ßen um eine Einheit resultieren. Die Log Odds geben die Chance für eine Realisierung des betrachteten Ereignisses im Vergleich zum Komplementärereignis wieder. Für sie hat sich die Bezeichnung "Logit" ein-gebürgert.

• Marginaler Effekt der erklärenden Variablen $x_j$ auf die abhängige Variable y

Die Einflüsse einer Veränderung der erklärenden Variablen auf die Wahrscheinlichkeit des betrachteten Ereignisses lassen sich dagegen durch die partiellen Ableitungen

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

angeben, in der f die Dichtefunktion der Logit-Verteilung bezeichnet:

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

Durch die Koeffizienten $\beta_j$ allein ist nur noch die Richtung determiniert, in die sich die Wahrscheinlichkeiten $p_t$ bei bestimmten Impulsen der exogenen Variablen verändern. Das Ausmaß der Veränderung hängt jedoch zusätzlich von der Steilheit der Verteilungsfunktion der Logit-Verteilung im Punkt $\boldsymbol{\beta}'\mathbf{x}_t$ ab. Je steiler die Verteilungsfunktion beim aktuellen Niveau verläuft, umso größer ist der Effekt einer Veränderung der erklärenden Variablen auf die Wahrscheinlichkeit des betrachteten Ereignisses.

● Maximum-Likelihood-Schätzung des Logit-Modells

$p_t$: Wahrscheinlichkeit, dass die Zufallsvariable $y_t$ den Wert 1 annimmt

bei Gültigkeit des Logit-Modells: $p_t = e^{\boldsymbol{\beta}'\mathbf{x}_t} / 1 + e^{\boldsymbol{\beta}'\mathbf{x}_t}$ (Gl. 7.15)

Vektor $\mathbf{x}_t$ der erklärenden Variablen: Merkmalsprofile der Individuen

→ Schätzung des Logit-Modells mit der **Maximum-Likelihood-Methode**

Prinzip der Maximum-Likelihood-Methode (ML-Methode):

Aus welcher Grundgesamtheit stammt eine konkrete Stichprobe mutmaßlich? Wenn man die Parameter der Grundgesamtheit variiert, wird sich auch die Wahrscheinlichkeit dafür ändern, dass sich die vorliegende Stichprobe realisiert. Unter welchen Werten der Parameter kommt aber der gegebenen Stichprobe die maximale Wahrscheinlichkeit zu, sich zu realisieren? Diese Werte heißen **Maximum-Likelihood-Schätzer (ML-Schätzer)**.

Stichprobe der Werte (0 oder 1) der abhängigen Variablen y: $y_1, y_2, ..., y_n$

Wahrscheinlichkeit, dass beim Individuum t $y_t$=1 oder $y_t$=0 eintritt:

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

für $y_t$=1: $f(y_t=1) = p_t$, für $y_t$=0: $f(y_t=0) = 1-p_t$

Gemeinsame Wahrscheinlichkeitsfunktion $f_y$ für $Y_1=y_1$, $Y_2=y_2$, ..., $Y_n=y_n$ bei einem Bernoulli-Prozess (unabhängige Ziehungen):

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

Likelihood-Funktion:

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

Zum Zwecke einer Schätzung des Parametervektors β betrachten wir die logarithmierte Likelihood-Funktion L*:

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

Bedingung erster Ordnung für ein Maximum der Likelihood-Funktion:

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

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

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

Da (7.19) ein nicht-lineares Gleichungssystem in $\boldsymbol{\beta}$ ist, kann der unbekannte Parametervektor allerdings nicht analytisch ermittelt werden. Vielmehr ist ein iteratives Verfahren anzuwenden, mit dem der Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\beta}}_{ML}$ für $\boldsymbol{\beta}$ numerisch bestimmt werden kann.

Ein solches Iterationsverfahren. Das eine hohe Konvergenzgeschwindigkeit besitzt, ist das Newton-Raphson-Verfahren.

Ausgehend von einem Anfangsschätzer $\boldsymbol{\beta}^{(0)}$ wird mit dem Newton-Raphson-Verfahren im m-ten Schritt ein verbesserter Parameterschätzer aus einem iterativen Schema der Form ermittelt:

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

Die Iterationsvorschrift basiert auf einer Taylor-Reihenzerlegung zweiter Ordnung. Für das Logit-Modell sind der Vektor der ersten Ableitungen (= Gradient) und die Matrix der zweiten Ableitungen (= Hesse-Matrix) durch

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

und

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

gegeben.

Die Hesse'sche Inverse kann zugleich als Schätzer für die Kovarianzmatrix des ML-Schätzers $\hat{\boldsymbol{\beta}}_{ML}$ verwendet werden. Ihre Diagonalelemente, die die Varianzen der Komponenten von $\hat{\boldsymbol{\beta}}_{ML}$ enthalten, lassen sich für einen Signifikanztest der geschätzten Koeffizienten des Logit-Modells verwenden.

**Beispiel:**

Logit-Modell für den Produktkauf (y) in Abhängigkeit vom Einkommen ($x_2$) und Geschlecht ($x_3$) lautet:

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

- Maximum-Likelihood-Schätzung (ML-Schätzung) des Logit-Modells

Teil 1: Regressionskoeffizienten mit Standardfehlern, z- und p-Werten

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 |

• Geschätztes Logit-Modell für den Produktkauf

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

Effekte der erklärenden Variablen Einkommen ($x_2$) und Geschlecht ($x_3$) auf die Kaufwahrscheinlichkeit p:

$\hat{\beta}_2 = 0,00448$ > 0: Einkommen ($x_2$) ↑ → Kaufwahrscheinlichkeit p ↑

$\hat{\beta}_3 = -2,87446$ < 0: Geschlecht ($x_3$) weibl. (1) statt männl. (0) → Kaufwahrsch. p ↑

Kaufwahrscheinlichkeit für den Konsumenten 1, der sich für den Nichtkauf ($y_1$=0) entschieden hat, aufgrund seines Merkmalsprofils $\mathbf{x}'_1$ = (1  2500  1):

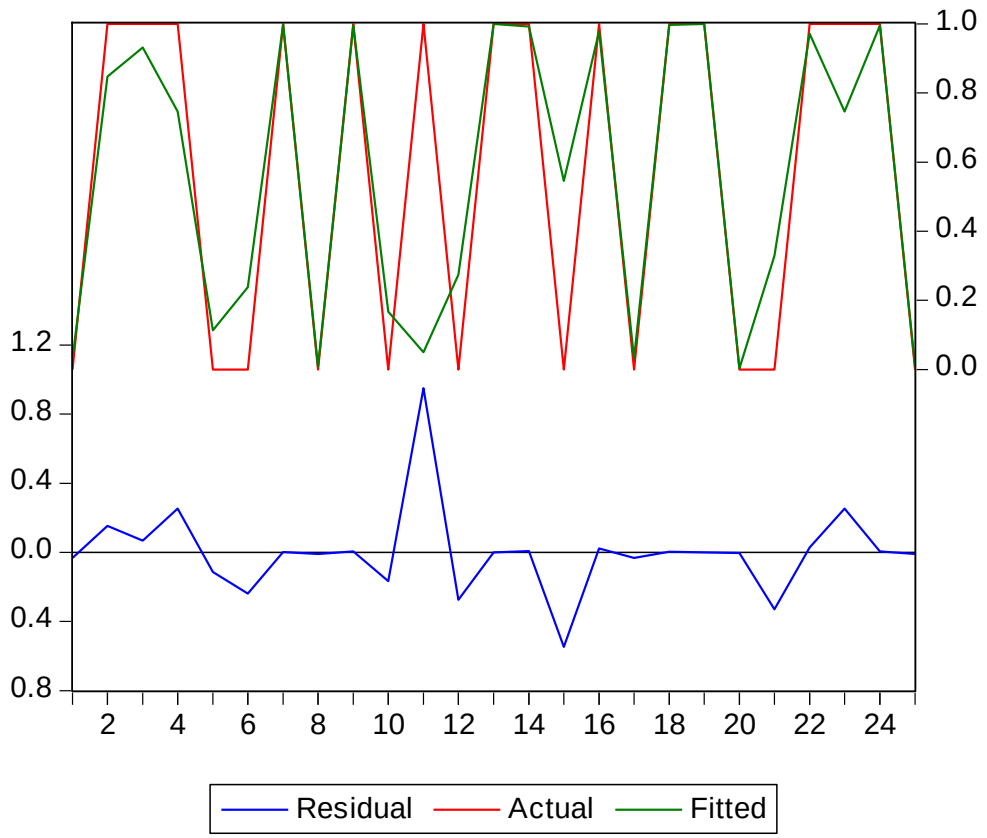
$$\hat{p}_1 = \frac{1}{1+e^{11,72644-0,00448\cdot 2500+2,87446\cdot 1}} = 0,0323$$

Kaufwahrscheinlichkeit für den Konsumenten 2, der sich für den Kauf ($y_2$=1) entschieden hat, aufgrund seines Merkmalsprofils $\mathbf{x}'_2$ = (1  3000  0):

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

- Beobachtungswerte der abh. Variablen, geschätzte Wahrscheinl., Residuen

| $t$ | $y_t$ | $\hat{p}_t$ | $\hat{u}_t$ |
|---|---|---|---|
| 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 |



• Varianz-Kovarianz-Matrix der Regressionskoeffizienten

- berechnet als Inverse der Matrix der zweiten Ableitungen (= Hesse‘sche Inverse)

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

(Hesse-Matrix an der Stelle $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{ML}$)

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

Standardfehler der Regressionskoeffizienten:

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

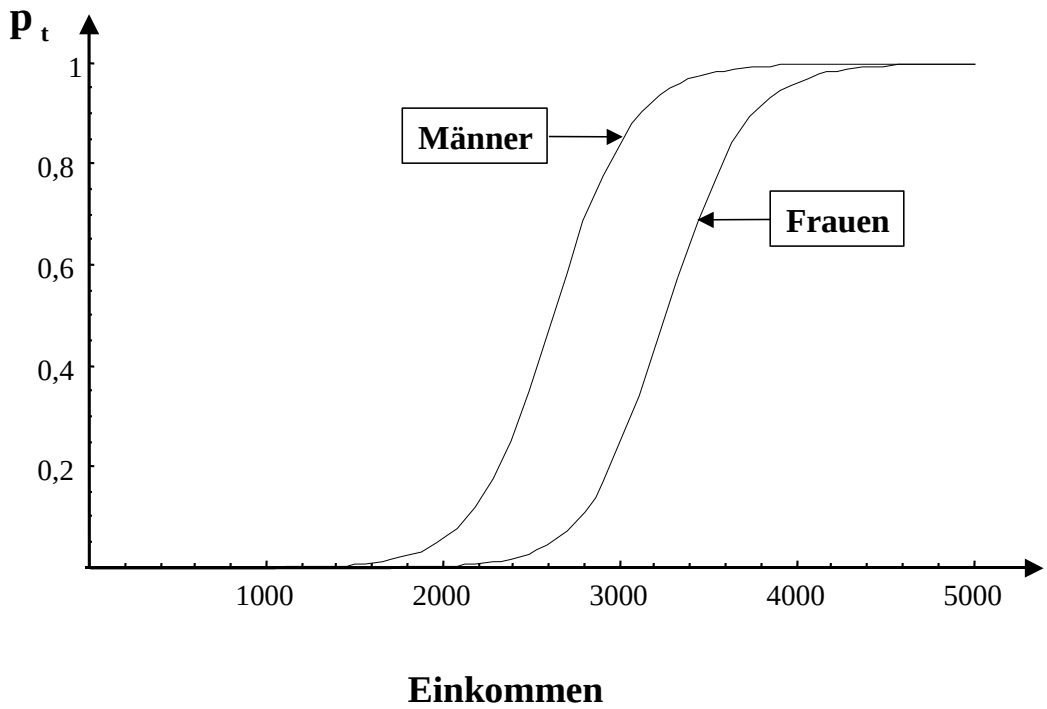
$$\hat{\sigma}_{\hat{\beta}_2} = \sqrt{0{,}00000416} = 0{,}0020240$$

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

- Verlauf der geschätzten kumulierten Kaufwahrscheinlichkeiten für Frauen und Männer in Abhängigkeit vom Einkommen ($x_2$)

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

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




Aus der Grafik wird aber auch deutlich, dass die Regressionskoeffizienten nicht genauso interpretiert werden können wie bei der linearen Regression. Wird das Einkommen im mittleren Bereich bei 3000€ um eine Einheit erhöht, dann steigt die Kaufwahrscheinlichkeit viel stärker als bei einer gleich starken Veränderung in den Randbereichen (beispielsweise bei 1000 € oder bei 5000 €).

• Marginaler Effekt des Einkommens ($x_2$) auf die Kaufwahrscheinlichkeit p

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

- Effekt einer Erhöhung des Einkommens um 100 € beim Konsumenten 1 (t=1) auf die Kaufwahrscheinlichkeit des Produkts:

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

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

- Effekt einer Erhöhung des Einkommens um 100 € beim Konsumenten 2 (t=2) auf die Kaufwahrscheinlichkeit des Produkts:

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

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

• Effektkoeffizienten

- Geschätzter Effektkoeffizient des Einkommens ($x_2$)

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

Mit zunehmendem Einkommen steigen die Odds, d.h. nimmt das Chancenverhältnis für den Produktkauf gegenüber dem Nichtkauf zu.

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

Erhöht sich das Einkommen um 1 GE, dann steigen die Odds innerhalb der Gruppe weiblicher und männlicher Konsumenten um den Faktor 1,0045, d.h. um 0,45%.

- Geschätzter Effektkoeffizient des Geschlechts ($x_3$)

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

Bei einem Übergang von der Gruppe der männlichen Konsumenten ($x_2 = 0$) zur Gruppe der weiblichen Konsumenten ($x_3 = 1$) sinken die Odds, d.h. nimmt das Chancenverhältnis für den Produktkauf gegenüber dem Nichtkauf ab.

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

Geht man von der Gruppe der männlichen Konsumenten ($x_2 = 0$) zur Gruppe der weiblichen Konsumenten ($x_3 = 1$) über, sinken die Odds bei Konstanthaltung des Einkommens um den Faktor 17,7, d.h. um 94,36%.

● Likelihood-Verhältnis-Test und Pseudo-$R^2$

Die Gültigkeit des Logit-Modells überhaupt lässt sich unter Verwendung des **Likelihood-Verhältnis-Testes** überprüfen. Wenn die Nullhypothese

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

gültig ist, kann die Wahrscheinlichkeit dafür, dass die Zufallsvariable y den Wert 1 annimmt, dadurch geschätzt werden, dass man die Anzahl $n_1$ der statistischen Einheiten, die die betrachtete Merkmalsausprägung haben, auf die Anzahl n der Beobachtungen bezieht:

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

Eine Bedingtheit durch exogene Variablen entfällt dann, da sie keinerlei Erklärungsgehalt besitzen. Unter der Nullhypothese ist die logarithmierte Likelihood-Funktion durch

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

gegeben. Während $L_0^*$ als restringierte log-Likelihood-Funktion bezeichnet wird, heißt die unter dem Logit-Modell resultierende logarithmische Likelihood-Funktion L* unrestringiert. Unter der Nullhypothese (7.20) folgt die Likelihood-Ratio-Statistik

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

asymptotisch einer $\chi^2$-Verteilung mit k-1 Freiheitsgraden. Eine Annahme der Nullhypothese impliziert, dass exogenen Variablen $x_2, x_3, \ldots, x_k$ keinen Einfluss auf die Wahrscheinlichkeit des betrachteten Ereignisses haben, so dass ihre Schätzung allein durch den Anteilswert $\hat{p}$ gerechtfertigt ist.

Als Maß für die Güte der Anpassung ist ein dem Determinationskoeffizienten vergleichbares Maß entwickelt worden, das aus diesem Grund als **Pseudo-$R^2$** bezeichnet wird

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

Das Pseudo-$R^2$ kann als Gütemaß verstanden werden, das den Grad der "Unsicherheit" in den Daten misst, der durch das Logit-Modell "erklärt" wird. Sofern $p_t$ stets 1 ist, wenn $y_t=1$ ist und 0, wenn $y_t=0$ ist, dann gilt für die unrestringierte Log-Likelihoodfunktion $L^*=\ln 1=0$, d.h. $PR^2$ erreicht dann seinen maximalen Wert. In diesem Extremfall leisten die erklärenden Variablen eine perfekte Modellanpassung Offenbar kann diese Situation jedoch nur dann eintreten, wenn $\boldsymbol{\beta}'\mathbf{x}_t$ gegen $-\infty$ und $+\infty$ geht. Der minimale Wert des Pseudo-$R^2$ wird genau dann erreicht, wenn die unrestringierte log-Likelihoodfunktion mit der restringierten übereinstimmt, d.h. wenn die exogenen Variablen keinerlei Erklärungsgehalt besitzen. Zwar kann grundsätzlich mit wachsendem $PR^2$ von einer verbesserten Modellanpassung gesprochen werden, doch besitzen die Werte des Pseudo-$R^2$ nicht die anschauliche Interpretation, die dem Determinantenkoeffizienten beim multiplen Regressionsmodell zukommt.

**Beispiel:**

Kenngrößen und Tests über die Modellanpassung für das geschätzte Logit-Modells:

| | | | |
|---|---|---|---|
| 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-$R^2$

  - Wert der Log Likelihood-Funktion unter der Nullhypothese (restringierte log Likelihood-Funktion $L_0^*$ ):

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

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

  - Wert der log Likelihood-Funktion für das geschätzte Logit-Modell (un-restringierte log Likelihood-Funktion L*):

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

Berechnung s. nächste Seite)

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

Berechnung der unrestringierten log Likelihood-Funktion 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 |
| $\Sigma$ | 14 | - | -3,9001 | -2,1718 | $L^* = -6,0719$ |

- Likelihood-Verhältnis-Test

Nullhypothese $H_0$: $\beta_2 = \beta_3 = 0$

Prüfgröße (Likelihood-Verhältnis-Statistik):

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

Kritischer Wert ($\alpha = 0{,}05$):

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

Testentscheidung:

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

- Klassifikationstabellen

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