

**Tests auf Heteroskedastie und
unterdrückte Variablen
im Ordered Probit-Modell**

Ruth Brand und Anja König*
Diskussionspapier 204
ISSN 0949-9962

Juni 1997

* Adresse der Autoren:
Institut für Quantitative Wirtschaftsforschung
Königsworther Platz 1
30167 Hannover
Fax-Nr. 0049 511 762-3923

1 Einleitung

Abhängige ordinal skalierte Variablen werden in den Wirtschafts- und Sozialwissenschaften in vielfältigen Zusammenhängen untersucht. Ein häufig für die Analyse genutzter Ansatz ist dabei das von McKelvey und Zavoina (1975) in die Literatur eingeführte Ordered Probit-Modell. Dabei liegt dem Zusammenhang für die beobachteten Variablen eine latente lineare Regressionsbeziehung zugrunde, deren Störterme unabhängig identisch normalverteilt sind. Die Verletzung der Annahmen bezüglich der Störterme führt ebenso wie die Unterdrückung einflußreicher Regressoren in diesem Ansatz zu inkonsistenten Parameterschätzungen. Aus diesem Grund werden in der neueren Literatur Spezifikationstests (Glewwe 1997) und Modifikationen des Ordered Probit-Modells (z.B. Kuckuck/Ronning 1996) diskutiert.

Im vorliegenden Beitrag wird der Ansatz von McKelvey und Zavoina erweitert um Heteroskedastie im latenten Modell (Abschnitt 2). Anschließend werden für das Standard-Modell Tests auf Heteroskedastie und unterdrückte Variablen entwickelt (Abschnitt 3), und ihre Eigenschaften werden in einer Simulationsstudie untersucht (Abschnitt 4).

2 Ordered Probit-Modell

Sei y_i^* , $i = 1, \dots, n$, eine nicht beobachtbare abhängige Variable, für die der lineare Regressionszusammenhang

$$y_i^* = x_i' \beta + u_i \tag{1}$$

gelte. Dabei bezeichnet x_i den $((K + 1) \times 1)$ -Vektor der beobachtbaren erklärenden Variablen mit der Konstanten $x_{0i} = 1$ und β einen $((K + 1) \times 1)$ -Vektor von unbekanntem Koeffizienten; u_i symbolisiert den Störterm des aufgrund der Nichtbeobachtbarkeit der endogenen Variablen latenten Modells (1). Für die Restgrößen $u_i, i = 1, \dots, n$, gelte, daß sie unabhängig über die n Erhebungseinheiten normalverteilt sind mit dem Erwartungswert 0 und der Varianz σ_i^2 , also $u_i \sim i.d. N(0, \sigma_i^2)$.

Beobachtet werden J Kategorien einer diskreten polytomen Indikatorvariablen mit ordinalem Meßniveau y_i , für die angenommen wird, daß sie zu y_i^* in folgendem monotonen Zusammenhang steht

$$y_i = \begin{cases} 1 & \text{wenn } \gamma_0 < y_i^* \leq \gamma_1 \\ 2 & \text{wenn } \gamma_1 < y_i^* \leq \gamma_2 \\ \vdots & \vdots \\ j & \text{wenn } \gamma_{j-1} < y_i^* \leq \gamma_j \\ \vdots & \vdots \\ J & \text{wenn } \gamma_{J-1} < y_i^* \leq \gamma_J \end{cases} . \quad (2)$$

$\gamma_1, \dots, \gamma_{J-1}$ stellen hier unbekanntes Schwellenwerte dar, bei deren Überschreiten die manifeste Variable y_i in eine andere Kategorie wechselt. Dabei gilt

$$\gamma_0 < \gamma_1 < \dots < \gamma_{J-1} < \gamma_J \quad ,$$

und da aufgrund der Normalverteilung des Störterms die latente Variable einen unbeschränkten Wertebereich besitzt, ferner $\gamma_0 = -\infty$ und $\gamma_J = \infty$.

Für die Wahrscheinlichkeit Kategorie j zu beobachten folgt aus dem Zu-

sammenhang (2)

$$\begin{aligned} P(y_i = j) &= P(\gamma_{j-1} < y_i^* \leq \gamma_j) \\ &= P(u_i \leq \gamma_j - x_i' \beta) - P(u_i \leq \gamma_{j-1} - x_i' \beta), \end{aligned}$$

und in Verbindung mit der Annahme $u_i \sim N(0, \sigma_i^2)$ ergibt sich

$$P(y_i = j) = \Phi\left(\frac{\gamma_j - x_i' \beta}{\sigma_i}\right) - \Phi\left(\frac{\gamma_{j-1} - x_i' \beta}{\sigma_i}\right), \quad (3)$$

wobei $\Phi(\cdot)$ die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Die Schätzung der unbekannt identifizierbaren Parameter des Modells erfolgt aufgrund der bei korrekter Modellspezifikation gegebenen Best Asymptotically Normal (BAN)-Eigenschaften der resultierenden Schätzfunktionen üblicherweise mittels der Maximum-Likelihood-Methode. Dabei ergibt sich die Likelihoodfunktion wegen der Unabhängigkeit der Beobachtungseinheiten als Produkt der n Einzelwahrscheinlichkeiten, für Beobachtung i Kategorie j zu erhalten:

$$L = \prod_{i=1}^n \prod_{j=1}^J \left[\Phi\left(\frac{\gamma_j - x_i' \beta}{\sigma_i}\right) - \Phi\left(\frac{\gamma_{j-1} - x_i' \beta}{\sigma_i}\right) \right]^{y_{ji}}. \quad (4)$$

Hier bezeichnet y_{ji} eine Dummy-Variable mit

$$y_{ji} = \begin{cases} 1 & \text{wenn } y_i = j \\ 0 & \text{sonst} \end{cases}$$

und dem Erwartungswert

$$E(y_{ji}) = 1 \cdot P(y_{ji} = 1) + 0 \cdot P(y_{ji} = 0) = P(y_i = j).$$

Die Dummy-Variable y_{ji} stellt sicher, daß das für jede Erhebungseinheit i gebildete Produkt über alle J Kategorien der manifesten endogenen Variablen

nur für die von diesem Beobachtungsträger realisierte Kategorie j den von Eins verschiedenen Faktor (3) enthält. Neben den $(K + 1)$ -Koeffizienten in β sind in (4) die n Standardabweichungen σ_i unbekannt, die als inzidentelle Parameter (Neyman/Scott 1948) in diesem Ansatz nicht konsistent schätzbar sind. Dieses Problem läßt sich mittels einer Parametrisierung lösen, in welcher die Heteroskedastie des Modells in funktionaler Abhängigkeit von einer oder mehreren Variablen modelliert wird. Hier stehen verschiedene Möglichkeiten zur Verfügung. So kann zum einen eine durch die abhängige Variable y_i^* ausgelöste Heteroskedastie angenommen werden, also $\sigma_i^2 = g(y_i^*)$ mit beliebiger Funktion g . Zum anderen läßt sich eine Abhängigkeit der individuen-spezifischen Varianz von einzelnen Regressoren und/oder anderen nicht im Regressionszusammenhang (1) berücksichtigten Variablen unterstellen.

Letzterem Ansatz soll hier mit der üblichen Formulierung

$$\sigma_i^2 = \sigma^2 \exp(2 z_i' \lambda) = \sigma^2 \exp\left(2 \sum_{l=1}^L z_{il} \lambda_l\right) \quad (5)$$

gefolgt werden. Dabei bezeichnet z_i einen $(L \times 1)$ -Vektor von beobachtbaren Variablen, und λ ist ein $(L \times 1)$ -Vektor unbekannter Koeffizienten. In dieser Modellierung wird die Heteroskedastie über den gemeinsamen Einfluß der z_{il} auf σ_i ausgelöst. Der Einfluß der Variablen z_{il} auf σ_i ist um so höher je größer λ_l ist. Für $\lambda = 0$ folgt $\sigma_i^2 = \sigma^2$ für alle $i = 1, \dots, n$ und damit das homoskedastische Standard Ordered Probit-Modell, das von McKelvey/Zavoina (1975) in die Literatur eingeführt wurde und durch die Beziehungen

$$\begin{aligned} u_i &\sim i.i.d. N(0, \sigma^2) \\ P(y_i = j) &= \Phi\left(\frac{\gamma_j - x_i' \beta}{\sigma}\right) - \Phi\left(\frac{\gamma_{j-1} - x_i' \beta}{\sigma}\right) \end{aligned} \quad (6)$$

gekennzeichnet ist. Für beliebiges λ gilt hingegen

$$P(y_i = j) = \Phi\left(\frac{\gamma_j - x'_i\beta}{\sigma \exp(z'_i\lambda)}\right) - \Phi\left(\frac{\gamma_{j-1} - x'_i\beta}{\sigma \exp(z'_i\lambda)}\right), \quad (7)$$

wobei $\Phi(\cdot)$ wiederum die Verteilungsfunktion der Standardnormalverteilung bezeichnet, und die Likelihoodfunktion (4) lautet dann

$$L = \prod_{i=1}^n \prod_{j=1}^J \left[\Phi\left(\frac{\gamma_j - x'_i\beta}{\sigma \exp(z'_i\lambda)}\right) - \Phi\left(\frac{\gamma_{j-1} - x'_i\beta}{\sigma \exp(z'_i\lambda)}\right) \right]^{y_{ji}}.$$

Sie enthält die $J + K + L + 1$ unbekanntes, zu schätzenden Parameter $\theta_0 = (\beta', \gamma', \lambda', \sigma)'$, von denen, sofern $x'_i\beta$ ein absolutes Glied enthält, wegen

$$\frac{\gamma_j - x'_i\beta}{\sigma \exp(z'_i\lambda)} = \frac{\gamma_j/\sigma - \beta_0/\sigma - \sum_{k=1}^K x'_i\beta_k/\sigma}{\exp(z'_i\lambda)}$$

nur $J + K + L - 1$ identifizierbar und damit konsistent schätzbar sind. Es sind also zwei Parameter zu normieren. Üblicherweise wird $\sigma = \gamma_1 = 1$ gesetzt. Alternativ kann auch die Normierung $\sigma = 1$ und $\beta_0 = 0$ gewählt werden, die das latente Modell (1) in eine homogene Regression überführt.

Die Maximierung der Likelihoodfunktion bzgl. θ unter der gewählten Normierung ergibt unter den üblichen Regularitätsbedingungen und bei korrekter Modellspezifikation einen konsistenten und asymptotisch effizienten Schätzer $\hat{\theta}$ mit

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, E\{-H(\theta_0)\}^{-1}\right), \quad H(\theta_0) = \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \Big|_{\theta=\theta_0}.$$

Die Eindeutigkeit der Schätzung folgt aus der globalen Konkavität der Likelihoodfunktion (Pratt 1981).

3 Tests

Während unter korrekter Modellspezifikation das gewählte Maximum-Likelihood-Verfahren BAN-Schätzer liefert, führen Fehlspezifikationen in der Regel zu inkonsistenten Schätzfunktionen (Ruud 1983). Falsche Parametrisierungen der Verteilungsannahme für die latente endogene Variable führen zu einer inkorrekten Formulierung der Wahrscheinlichkeiten (3) - vgl. auch (6) und (7) - und damit zur Fehlspezifikation der Likelihoodfunktion. Die resultierenden Schätzfunktionen sind damit zwar konsistent im Hinblick auf den Pseudo-True-Value θ_* , der das Kullback-Leibler-Informations-Kriterium zwischen der wahren und der fehlspezifizierten Verteilung minimiert (Akaike 1973, White 1982), in der Regel sind sie aber inkonsistent hinsichtlich der interessierenden Parameter θ_0 . Das gilt für die Schätzung eines homoskedastischen Modells unter Heteroskedastie der unbeobachteten Variablen ebenso wie für die Schätzung bei Unterdrückung einflußreicher Variablen in der latenten Regressionsbeziehung, im letzteren Fall also für die Schätzung des Modells

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

statt eines korrekten vollständigen Modells

$$y_i^* = x_i' \beta + w_i' \alpha + \varepsilon_i \quad , \quad (8)$$

in dem w_i den $(M \times 1)$ -Vektor der nicht berücksichtigten Regressoren und α den $(M \times 1)$ -Vektor der dazugehörigen Koeffizienten symbolisiert.

Damit ist es notwendig, die Validität der Modellspezifikation mittels Tests zu überprüfen, die auf Basis der asymptotischen Normalverteilung der ML-

Schätzer bei korrekter Spezifikation über die klassischen Testprinzipien entwickelt werden können. Im folgenden werden ein Test auf Heteroskedastie, der ausgehend von der Formulierung (5) die Hypothese

$$H_0 : \lambda = 0 \quad \text{gegen} \quad H_1 : \lambda \neq 0 \quad (9)$$

und damit des homoskedastische Standard Ordered Probit-Modell gegen das allgemeinere heteroskedastische Modell testet, und ein Test auf unterdrückte Variablen vorgestellt; letzterer prüft die Parameterrestriktion

$$H_0 : \alpha = 0 \quad \text{gegen} \quad H_1 : \alpha \neq 0 \quad (10)$$

in der latenten Regressionsbeziehung (8), wobei der Test für das homoskedastische Standard Ordered Probit-Modell formuliert wird¹. Die Teststatistiken werden jeweils auf Basis des Likelihood-Ratio-Prinzips, des Wald-Prinzips und des Lagrange-Multiplikator-Prinzips entwickelt (z. B. Engle 1984):

Der Likelihood-Ratio-Test basiert auf dem Vergleich der Maxima der Likelihoodfunktion des Modells unter der Null- und der Gegenhypothese. Die allgemeine Likelihood-Ratio-Teststatistik lautet

$$LR = 2(\ln L_1 - \ln L_0) \overset{a}{\sim} \chi_G^2 \quad ,$$

wobei mit $\ln L_1$ das Maximum der logarithmierten Likelihoodfunktion des unrestringierten und mit $\ln L_0$ das der Loglikelihoodfunktion des restringierten Ansatzes bezeichnet wird.

¹Die Teststatistik läßt sich problemlos auf das heteroskedastische Modell verallgemeinern.

Der Wald-Test vergleicht die Schätzwerte für die Modellparameter des unrestringierten Modells mit den in H_0 spezifizierten restringierten Parametervektor. Für die hier formulierten Testprobleme lautet die Teststatistik

$$W = \widehat{\tilde{\theta}}_{H_1}' V(\widehat{\tilde{\theta}}_{H_1})^{-1} \widehat{\tilde{\theta}}_{H_1} \stackrel{a}{\sim} \chi_G^2 \quad ,$$

wobei $\widehat{\tilde{\theta}}_{H_1}$ die Schätzfunktionen $\hat{\lambda}$ im Test auf Heteroskedastie bzw. $\hat{\alpha}$ für den Test auf unterdrückte Variablen bezeichnet, beides geschätzt im unrestringierten Modell, und $V(\widehat{\tilde{\theta}}_{H_1})$ ist die Varianz dieser unrestringierten Schätzung.

Der Lagrange-Multiplikator-Test basiert auf den Steigungen der logarithmierten Likelihoodfunktion des unrestringierten Modells, berechnet für die Schätzwerte des restringierten Ansatzes. Sind die in der Nullhypothese spezifizierten Restriktionen zutreffend, ist der Erwartungswert des betrachteten Gradienten der Loglikelihoodfunktionen der Nullvektor. Die Teststatistik lautet

$$LM = \nabla(\hat{\theta}_{H_0})' V(\hat{\theta}_{H_0})^{-1} \nabla(\hat{\theta}_{H_0}) \stackrel{a}{\sim} \chi_G^2 \quad ,$$

wobei $\hat{\theta}_{H_0}$ den unter den formulierten Parameterrestriktionen geschätzten Parametervektor bezeichnet, und $\nabla(\hat{\theta}_{H_0})$ ist der Vektor der ersten Ableitungen des allgemeinen Modells an der Stelle $\hat{\theta}_{H_0}$. $V(\hat{\theta}_{H_0})$ bezeichnet die Inverse der unter H_0 bestimmten Varianz-Kovarianzmatrix des Scorevektors $\nabla(\hat{\theta}_{H_0})$; diese Inverse ist identisch zu der unter H_0 gebildeten Varianz-Kovarianzmatrix des Schätzers $\hat{\theta}_{H_0}$.

Eine ausführliche Darstellung der Wald- und LM-Teststatistiken findet sich in den Abschnitten D und E des Anhangs, die für die Herleitung der

Tests relevanten partiellen Ableitungen erster und zweiter Ordnung der Loglikelihoodfunktionen der Modelle in den Abschnitten B und C.

Die drei vorgestellten Teststatistiken sind unter der Nullhypothese asymptotisch äquivalent und jeweils χ^2 -verteilt mit G Freiheitsgraden, wobei die Zahl der Freiheitsgrade der Zahl der Parameterrestriktionen entspricht, also $G = L$ für den Test auf Heteroskedastie und $G = M$ für den Test auf unterdrückte Variablen. Die Nullhypothese wird bei allen drei Tests abgelehnt, wenn der Wert der Teststatistik den aus der χ_G^2 -Verteilung bestimmten kritischen Wert übersteigt. In kleinen Stichproben kann die Verteilung der drei Statistiken erheblich von der asymptotischen Verteilung abweichen.

4 Simulationen

Das Verhalten der Tests auf Heteroskedastie und unterdrückte Variablen wurde in umfangreichen Simulationen untersucht. Die Berechnung der Teststatistiken setzt die Maximierung der Likelihoodfunktion bzw. der Loglikelihoodfunktion des restringierten oder des unrestringierten Modells voraus. Aufgrund der Nichtlinearität der ersten Ableitungen in beiden Ansätzen ist dieses nur numerisch möglich. Üblicherweise wird hier das Newton-Raphson-Verfahren verwendet, bei dem $\hat{\theta}$ iterativ aus

$$\hat{\theta}^s = \hat{\theta}^{s-1} - [H(\hat{\theta}^{s-1})]^{-1} \nabla(\hat{\theta}^{s-1}) \quad (11)$$

bestimmt wird, wobei $H(\hat{\theta}^{s-1})$ die Hessesche Matrix $\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \Big|_{\theta = \hat{\theta}^{s-1}}$ und $\nabla(\hat{\theta}^{s-1})$ den Gradienten $\frac{\partial \ln L}{\partial \theta} \Big|_{\theta = \hat{\theta}^{s-1}}$ der Loglikelihoodfunktion bezeichnet, jeweils be-

rechnet an der Stelle $\hat{\theta}^{s-1}$. Alternativ können Modifikationen dieses Verfahrens genutzt werden, so die Scoring-Methode, bei der in (11) statt der Hesseschen Matrix ihr Erwartungswert, d. h. die Informationsmatrix, verwendet wird, oder der BHHH-Algorithmus (nach Berndt et al. 1974), der statt der Hesseschen Matrix das äußere Gradientenprodukt nutzt.

Die drei beschriebenen Verfahren konvergieren in einfachen Modellen wie dem binären Probit-Modell, das mit nur zwei Ausprägungen für die manifeste endogene Variable und einem Schwellenwert einen Spezialfall des Ordered Probit-Modells darstellt, in wenigen Iterationsschritten. In komplexeren Modellen sind die Konvergenzeigenschaften nicht befriedigend. Die Zahl der benötigten Iterationen ist sehr hoch, dementsprechend auch unter leistungsfähigen Computern der zeitliche Rechenaufwand. Wesentlich schneller konvergiert hier das Rang-Zwei-Verfahren, das definiert ist durch die Iteration

$$\hat{\theta}^s = \hat{\theta}^{s-1} - A^{s-1} \nabla (\hat{\theta}^{s-1})$$

mit

$$\begin{aligned} A^{s-1} &= \nu A^{s-2} + \left(1 + \nu \rho \frac{q' A^{s-2} q}{p' q}\right) \frac{pp'}{p' q} - \nu \frac{1 - \rho}{q' A^{s-2} q} A^{s-2} q q' A^{s-2} \\ &\quad - \frac{\nu \rho}{p' q} (p q' A^{s-2} + A^{s-2} q p') \quad , \quad \nu > 0, \rho \geq 0, \\ q &= \nabla (\hat{\theta}^{s-1}) - \nabla (\hat{\theta}^{s-2}) \\ p &= \hat{\theta}^{s-1} - \hat{\theta}^{s-2} \end{aligned}$$

und beliebiger Startmatrix A^0 , die positiv definit sein sollte, um sicherzustellen, daß ausgehend von einem Startvektor $\hat{\theta}^0$ der Gradient des ersten Schrittes

eine Abstiegsrichtung aufweist. Das Verfahren besitzt auch den rechentechnischen Vorteil, daß in den einzelnen Iterationen weder die Hessesche Matrix, noch die Informationsmatrix oder das äußere Gradientenprodukt bestimmt werden müssen.

In allen Simulationen wurde das beschriebene Rang-Zwei-Verfahren genutzt und zwar in der „Davidon/Fletcher/Powell (DFP)-Version“ mit $\nu = 1$ und $\rho = 0$ (benannt nach Davidon 1959 und Fletcher/Powell 1963, zitiert nach Fletcher/Powell 1963 sowie Stoer 1994).

Simuliert wurden vier Modelle, die detailliert in Tabelle 1 beschrieben sind. Das Grundmodell M.1 besteht aus einer homogenen linearen latenten Regressionsbeziehung mit vier exogenen Regressoren und einer normalverteilten homoskedastischen Störgröße sowie einer manifesten endogenen Variablen mit fünf Kategorien. Für das zweite Modell M.2 wird das Grundmodell um Heteroskedastie der Art (5) in der Restgröße des latenten Modells erweitert, wobei nur eine Einflußgröße z_i unterstellt wird, die außerdem nicht in der latenten Regressionsbeziehung berücksichtigt ist. Der dritte Ansatz M.3 entsteht aus einer Erweiterung des Grundmodells um eine weitere Einflußgröße w_i in der latenten Regressionsbeziehung. Das vierte Modell M.4 wird wie das Modell M.1 erzeugt, jedoch mit einer asymmetrischen Verteilung der Regressoren und anderen Koeffizienten und Schwellenwerten.

Für jedes dieser Modelle wurden die bereits beschriebenen Tests betrachtet. Zum einen wurde unter der Annahme einer korrekt spezifizierten latenten Regressionsbeziehung die Hypothese der Homoskedastie getestet, d. h., ausgehend von einem allgemeinen Modell mit Heteroskedastie und erklärenden

Tabelle 1: Spezifikationen

	M.1	M.2	M.3	M.4
y_i^*	$\beta_0 + \sum_{k=1}^4 x_{ki}\beta_k + u_i$	$\beta_0 + \sum_{k=1}^4 x_{ki}\beta_k + w_i\delta + u_i$	$\beta_0 + \sum_{k=1}^4 x_{ki}\beta_k + w_i\delta + u_i$	$\beta_0 + \sum_{k=1}^4 x_{ki}\beta_k + u_i$
x_i	$N(0, I_4), I = \text{Einheitsmatrix}$			$\text{Log}N(0, 0.5 \cdot I_4)$
z_i	-	$N(0, 1)$	-	-
w_i	-	-	$N(0, 1)$	-
u_i	$N(0, 1)$	$N(0, \exp(z_i \lambda))$	$N(0, 1)$	$N(0, 1)$
β	$(0, 1, 0.2, 0.05, 0.15)'$			$(0, 1, 0.5, 0.2, 0.4)'$
γ	$(-1.5, -0.5, 0.5, 1.5)'$			$(1.3, 1.9, 2.4, 3.05)'$
λ	-	0.05	-	-
δ	-	-	0.1	-
$y_i \in$	$\{1, 2, 3, 4, 5\}$			

Variablen (x'_i, w'_i) wurde ein Test auf die Parameterrestriktion

$$H_0 : \lambda = 0, \alpha = 0 \quad \text{gegen} \quad H_1 : \lambda \neq 0, \alpha = 0$$

durchgeföhrt. Zum anderen wurde ein Test auf unterdrückte Variablen unter der Annahme eines homoskedastischen Modells betrachtet, der, wiederum ausgehend von einer allgemeinen Modellformulierung, die Nullhypothese

$$H_0 : \lambda = 0, \alpha = 0 \quad \text{gegen} \quad H_1 : \lambda = 0, \alpha \neq 0$$

testet.²

²Die von den Darstellungen (9) und (10) abweichenden Hypothesenformulierungen für

Die Berechnung der Wald- und Lagrange-Multiplikator-Teststatistiken setzt eine Schätzung der Varianz-Kovarianzmatrix der geschätzten Parameter voraus. Hier wurden alternativ vier verschiedene Verfahren verwendet: Ein unter korrekter Modellspezifikation konsistenter Schätzer für die Varianz-Kovarianzmatrix der Parameterschätzungen ist die negative Inverse der Hesseschen Matrix, berechnet an der Stelle $\hat{\theta}$

$$\hat{V}_H = - \{H(\hat{\theta})\}^{-1} .$$

Gleiches gilt für die Inverse der Informationsmatrix

$$\hat{V}_I = \{E(-H(\hat{\theta}))\}^{-1} = I(\hat{\theta})^{-1}$$

und für die Inverse des äußeren Gradientenprodukts

$$\hat{V}_{OPG} = \left\{ \sum_{i=1}^n \frac{\partial \ln L(x_i, \hat{\theta})}{\partial \theta} \frac{\partial \ln L(x_i, \hat{\theta})}{\partial \theta'} \right\}^{-1} .$$

Die drei Schätzer sind asymptotisch äquivalent, können sich aber in kleinen Stichproben erheblich unterscheiden. Alternativ wurde der Schätzer

$$\hat{V}_W = \{-H(\hat{\theta})\}^{-1} \left\{ \sum_{i=1}^n \frac{\partial \ln L(x_i, \hat{\theta})}{\partial \theta} \frac{\partial \ln L(x_i, \hat{\theta})}{\partial \theta'} \right\} \{-H(\hat{\theta})\}^{-1}$$

nach White (1982,1983) verwendet, der auch bei Fehlspezifikationen des Modells die Varianz-Kovarianzmatrix der Parameterschätzungen konsistent schätzt, sofern die Stichprobenvariablen unabhängig identisch verteilt sind, und damit Tests im Hinblick auf den Pseudo-True-Value θ_* zum nominellen Testniveau erlaubt. Bei nicht identisch verteilten Stichprobenvariablen, wie

 beide Tests ergeben sich aus der expliziten Berücksichtigung der jeweils nicht getesteten anderen Form der Fehlspezifikation.

im Fall fixer Regressoren und/oder fixer individuenspezifischer Einflußgrößen in der Varianz der Störterme, erlaubt eine mit \hat{V}_W formulierte Teststatistik konservative Schlüsse (White 1983).

Insgesamt wurden damit die in der Tabelle 2 dargestellten Teststatistiken betrachtet und ausgewertet.

Tabelle 2: Berechnete Teststatistiken

Testart	Bezeichnung	Schätzung von $V = E \left(-\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \Big _{\theta=\theta_0} \right)^{-1}$
Lagrange-Multiplikator-Test	LM1	\hat{V}_I
	LM2	\hat{V}_{OPG}
	LM3	\hat{V}_H
Wald-Test	W1	\hat{V}_I
	W2	\hat{V}_{OPG}
	W3	\hat{V}_H
	W4	\hat{V}_W
Likelihood-Ratio-Test	LR	-

Die Modelle M.1 – M.4 wurden zweimal mit jeweils 5000 Replikationen simuliert. Dabei wurde jeweils eine feste Regressormatrix vorgegeben, so daß in jeder Wiederholung lediglich der Störterm und die endogene Variable neu erzeugt wurden. In den Tabellen A1 – A8 im Anhang A sind die Ergebnisse zusammengefaßt dargestellt, wobei die Ablehnraten der Tests bei einem

nominellen Signifikanzniveau von $\alpha = 5\%$ ausgewertet wurden. Alle Simulationen sowie die Berechnungen der Schätzer und der Teststatistiken wurden mit GAUSS 3.1.4 durchgeführt.

Generell ist anzumerken, daß sich die Berechnungen der Schätzer $\hat{\theta}$ und \hat{V} in kleinen Stichproben als numerisch relativ instabil erweisen, da die Werte der Dichte- und der Verteilungsfunktion der Normalverteilung über eine numerische Approximation bestimmt werden müssen. Aufgrund der Komplexität der Modelle werden hierdurch numerische Fehler induziert, die in sehr kleinen Stichproben auch zum Abbruch des Iterations- und damit des Schätzverfahrens führen können.

Aus diesem Grund wurden Simulationen für die Modelle M.1 – M.3 erst ab einem Stichprobenumfang von $n = 150$ durchgeführt, Modell 4 konnte sogar erst ab einem Stichprobenumfang von $n = 200$ stabil simuliert werden; die nicht symmetrische Verteilung der Regressoren führte zu vermehrten numerischen Problemen.

Die Ergebnisse zeigen, daß – wie zu erwarten – das nominale Testniveau unter H_0 von allen Tests – bis auf W4, den Wald-Test in der Variante von White (1982) – relativ schnell erreicht wird. Gemessen an einem Vergleich einer Kerndichteschätzung der Verteilung der Teststatistiken mit der Dichtefunktion der χ_1^2 -Verteilung läßt sich ferner feststellen, daß die Verteilung der Teststatistiken auch in kleinen Stichproben bereits gut durch die theoretische Verteilung approximiert werden kann. Exemplarisch findet sich eine Darstellung für den typischen Fall des Likelihood-Ratio-Tests auf Heteroskedastie in Modell M.1 für ausgewählte Stichprobenumfänge in Abb. 1 im

Anhang A. Beide Ergebnisse gelten sowohl für die Simulationen mit normalverteilten Regressoren (Tabellen A1 und A3) als auch für die Simulationen mit lognormalverteilten exogenen Variablen (Tabellen A7 und A8). Diese Ergebnisse weisen insgesamt darauf hin, daß die Konvergenzrate der Tests nur unwesentlich von der Symmetrie der Verteilung der Regressoren abhängt.

Ebenso streben alle Tests mit steigender Fallzahl unter ihrer Gegenhypothese gegen eine Ablehnrate von 1 (Tabellen A2 und A4). Dieses zeigt sich für alle üblichen Testniveaus, wie in Abb. 2 im Anhang A exemplarisch für den LR-Test auf Heteroskedastie gezeigt wird. Korrigiert man für den 'size-bias' in dem Sinne, daß statt des asymptotischen kritischen Wertes der χ_1^2 -Verteilung der kritische Wert aus den Simulationen unter H_0 (M.1) bestimmt wird, erhöht sich die Güte für kleine Stichproben im relevanten Bereich erheblich (vergl. Abb. 3, Anhang A). Darüber hinaus scheint die Güte des Tests auf Heteroskedastie vergleichsweise höher zu sein, die Ablehnraten übersteigen die des Tests auf unterdrückte Variablen für vergleichbare Stichprobenumfänge durchweg.

Ein weiteres zentrales Ergebnis ist, daß die Tests unberührt davon zu sein scheinen, ob die jeweils nicht getestete Form der Fehlspezifikation vorliegt. Der Test auf Heteroskedastie erreicht unter M.3 ebenso das nominale Testniveau wie der Test auf unterdrückte Variablen unter M.2 (Tabellen A5 und A6), wobei die Konvergenz nur unwesentlich langsamer ist als unter H_0 . Dieses Ergebnis wurde auch von Taylor (1991) für die LM2- und LM3-Statistiken des Tests auf unterdrückte Variablen im Tobit-Modell erzielt. Taylor stellt darüber hinaus fest, daß die Güte dieser Tests gegen ihre eigene Alternative

bei tatsächlicher Fehlspezifikation in Form von heteroskedastischen Störtermen und fehlenden Einflußgrößen in der latenten Regressionsbeziehung geringer ist als in den Simulationen unter der 'unverfälschten' Gegenhypothese, d. h. unter einer Fehlspezifikation, die analog zu M.3 nur durch die Unterdrückung von Variablen ausgelöst ist. Entsprechende Untersuchungen für das Ordered Probit-Modell stehen noch aus.

Für den Wald-Test nach White (W4) wird unter der hier gewählten Simulationsart mit einer fixen Regressormatrix erwartet, daß – wie oben ausgeführt – bei der Einführung einer zusätzlichen nicht getesteten Form der Fehlspezifikation ein konservativer Schluß gezogen wird. Dem entsprechend sollte der Test gegen ein nominales Testniveau von weniger als 5% konvergieren. Dieses ist allerdings in den durchgeführten Monte-Carlo-Experimenten nicht der Fall. Der Test konvergiert unter H_0 wesentlich langsamer als die zum Vergleich herangezogenen Statistiken gegen das nominale Testniveau und er unterschreitet es auch bei einem Stichprobenumfang von $n = 750$ nicht (Tabellen A5 und A6).

Darüber hinaus weist W4 in der Regel zwar höhere Ablehnraten unter der Gegenhypothese auf, sofern man den üblichen kritischen Wert aus der χ_1^2 -Verteilung heranzieht (Tabellen A2 und A4). Korrigiert man aber für den 'size-bias', ist die Güte der Teststatistik W4 nicht mehr höher als die der vergleichbaren Teststatistiken, wie in Abb. 4 im Anhang A exemplarisch für den Test auf Heteroskedastie bei einem Stichprobenumfang von $n = 150$ gezeigt wird. Insgesamt besitzt der Test damit in den hier durchgeführten Simulationen keine numerischen Vorteile.

Der Vergleich der Statistiken LM1 – LM3, W1 – W3 und der LR-Teststatistiken zeigt zunächst kein eindeutiges Ergebnis, da die verschiedenen Varianten der LM-Tests und der Wald-Tests sich in ihren Ergebnissen zum Teil erheblich unterscheiden. Jedoch kann tendenziell festgestellt werden, daß in beiden Tests die Ablehnraten der LM1-Statistik sich dem nominalen Testniveau geringfügig schneller als die des LR-Tests nähern und das dieser wiederum etwas schneller als die Wald-Statistik W2 konvergiert. Ein Vergleich der Statistiken W2 und W3 und der Statistiken LM2 und LM3 ergibt sowohl für den Test auf Heteroskedastie als auch für den Test auf unterdrückte Variablen keine eindeutige Reihung. Vielmehr bestätigt sich beim Vergleich der LM-Teststatistiken untereinander das auch aus anderen Monte-Carlo Experimenten bekannte Ergebnis, daß der unter der Verwendung der Informationsmatrix bestimmte LM-Test (LM1) unter der Nullhypothese durchweg besser ist als die Teststatistiken LM2 und LM3 (z.B. Davidson/MacKinnon 1983, 1984). Darüber hinaus zeigt sich beim Vergleich der Wald-Tests, daß die unter Verwendung der Hesseschen Matrix bestimmte Teststatistik W2 unter der Nullhypothese stets günstigere Eigenschaften aufweist als die Statistiken W1 und W3.

5 Fazit

Das Ordered-Probit-Modell ist ein häufig genutzter Ansatz zur Analyse ordinaler abhängiger Variablen. Dabei ist die Konsistenz der Schätzung der interessierenden Parameter abhängig von der exakten Spezifikation des Modells.

Eine durch nicht berücksichtigte Heteroskedastie hervorgerufene Fehlspezifikation führt ebenso wie die Unterdrückung von Variablen zu inkonsistenten Schätzungen für den Koeffizientenvektor.

Im vorliegenden Beitrag wird gezeigt, daß die Berücksichtigung von Heteroskedastie der Form $\sigma_i^2 = \sigma^2 \exp(2z_i'\lambda)$, wobei z_i einen Vektor beliebiger Regressoren beschreibt, im Ordered Probit-Modell möglich ist. Darüber hinaus werden auf Basis der klassischen Testprinzipien Tests auf Heteroskedastie und Tests auf unterdrückte Variablen entwickelt.

In den Simulationen zeigten sich für alle Tests – bis auf den Wald-Test nach White (W4) – die zu erwartenden Ergebnisse. Insbesondere konvergieren die Teststatistiken LM1 – LM3, W1 – W3 und die LR-Tests auch dann gegen ihr nominales Testniveau, wenn eine weitere Form der Fehlspezifikation vorliegt. Dabei sind die numerischen Unterschiede in den Ablehnraten relativ gering, so daß alle Teststatistiken geeignet erscheinen.

Die Teststatistik W4 weist gegenüber allen anderen Statistiken in kleinen Stichproben erhebliche numerische Nachteile auf. Dieses Ergebnis zeigt sich insbesondere auch, wenn zusätzlich eine nicht getestete Form der Fehlspezifikation in das Modell eingeführt wird.

Ein detaillierter Vergleich der einzelnen Teststatistiken zeigt, daß die Berechnung eines LM-Tests mit Hilfe der Informationsmatrix und die Berechnung eines Wald-Tests unter Verwendung der Hesseschen Matrix jeweils am geeignetsten erscheinen, da sie am schnellsten gegen das nominale Testniveau konvergieren und eine hohe Güte aufweisen.

Eine detailliertere Untersuchung der Güte der Tests ist anhand der vorliegenden Simulationsstudie nicht möglich. Insbesondere wurde das Verhalten der Tests, wenn sowohl Heteroskedastie als auch unterdrückte Variablen zu Fehlspezifikation führen, nicht untersucht. Darüber hinaus würde eine detailliertere Untersuchung der Güte der Tests Studien erfordern, die auch die Auswirkungen anderer zusätzlich eingeführter Formen der Fehlspezifikation, z.B. Abweichungen von der Normalverteilungsannahme, einbeziehen.

Literatur

- Akaike, H. 1973:** Information Theory and an Extension of the Likelihood Principle, in: Petrov, B.N. und F. Csaki, Proceedings of the Second International Symposium of Information Theory, Budapest 1973.
- Berndt, E.R., Hall, B.H., Hall, R.E. und J.A. Hausman 1974:** Estimation and Inference in non-linear structural models, *Annals of Economic and Social Measurement* 3, 653 – 656.
- Cramer, J.S. 1986:** *Econometric Applications of Maximum Likelihood methods*, Cambridge: University Press.
- Davidson, R. und J.G. MacKinnon 1983:** Small Sample Properties of Alternative Forms of the Lagrange Multiplier test, *Economics Letters* 12, 269 – 275.
- Davidson, R. und J.G. MacKinnon 1984:** Convenient Specification Tests for Logit and Probit Models, *Journal of Econometrics* 25, 241 – 262.
- Engle, R.F. 1984:** Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, in: Griliches, Z. und M.D. Intriligator, Hrsg., *Handbook of Econometrics*, Vol. II, Amsterdam: North Holland, S. 775-826.
- Fletcher, R. und M.J.D. Powell 1963:** A rapidly convergent descent method for minimization, *Computer Journal* 6, 163 – 168.
- Glewwe, P. 1997:** A Test of the Normality Assumption in the Ordered Probit Model, *Econometric Reviews* 16, 1 – 19.

- Kuckuck, M. und G. Ronning 1996:** Efficient Estimation of Ordered Probit Models, *Journal of the American Statistical Association* 91, 1120 – 1129.
- McKelvey, R.D. und W. Zavoina 1975:** A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology* 4, 103 – 120.
- Neymann, und Scott, 1948:** Consistent Estimates Based on Partially Consistent Observations, *Econometrica* 16, 1-32.
- Pratt, J.W. 1981:** Concavity of the Log Likelihood, *Journal of the American Statistical Association* 76, 103 – 106.
- Ruud, P.A. 1983:** Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models, *Econometrica* 51, 225-228.
- Stoer, J. 1994:** *Numerische Mathematik* 1, 7. Auflage, Berlin/Heidelberg u.a.O. 1994.
- Taylor, L.W. 1991:** Testing exclusion restrictions for a misspecified Tobit model, *Economics Letters* 37, 411 - 416.
- White, H. 1982:** Maximum Likelihood Estimation of Misspecified Models, *Econometrica* 50, 1 – 25.
- White, H. 1983:** Corrigendum, *Econometrica* 51, 513.

Anhang

A Ergebnisse

A.1 Simulationsergebnisse

Die Tabellen A1 bis A8 enthalten die Ablehnraten der Tests bei einem nominellen Testniveau von $\alpha = 5\%$, berechnet aus zwei Simulationsläufen mit je 5000 Wiederholungen.

Tabelle A1: Spezifikation M.1 – Test auf Heteroskedastie

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	6	0.049	0.078	0.107	0.077	0.063	0.068	0.117	0.060
200	0	0.047	0.069	0.085	0.068	0.058	0.063	0.094	0.055
250	0	0.051	0.069	0.079	0.068	0.060	0.062	0.088	0.058
300	0	0.052	0.068	0.072	0.063	0.056	0.061	0.079	0.057
400	0	0.048	0.060	0.064	0.057	0.054	0.055	0.071	0.052
500	0	0.048	0.059	0.060	0.055	0.052	0.053	0.066	0.051
750	0	0.049	0.054	0.057	0.055	0.053	0.053	0.060	0.051

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A2: Spezifikation M.2 – Test auf Heteroskedastie

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	9	0.259	0.309	0.354	0.304	0.274	0.294	0.371	0.280
200	0	0.301	0.361	0.397	0.358	0.325	0.346	0.413	0.332
250	0	0.390	0.435	0.466	0.435	0.401	0.424	0.482	0.414
300	0	0.432	0.481	0.509	0.481	0.451	0.473	0.520	0.461
400	0	0.594	0.626	0.646	0.627	0.605	0.621	0.655	0.614
500	0	0.653	0.672	0.685	0.676	0.660	0.672	0.695	0.665
750	0	0.873	0.886	0.889	0.885	0.874	0.881	0.891	0.880
1500	0	0.990	0.990	0.990	0.990	0.989	0.990	0.990	0.990

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A3: Spezifikation M.1 – Test auf unterdrückte Variablen

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	6	0.060	0.075	0.062	0.070	0.067	0.070	0.084	0.062
200	0	0.051	0.062	0.053	0.058	0.058	0.058	0.068	0.053
250	0	0.052	0.061	0.053	0.055	0.054	0.055	0.064	0.053
300	0	0.051	0.059	0.051	0.055	0.054	0.054	0.062	0.052
400	0	0.051	0.056	0.051	0.052	0.053	0.053	0.058	0.051
500	0	0.048	0.051	0.048	0.050	0.050	0.049	0.053	0.049
750	0	0.053	0.056	0.053	0.055	0.054	0.055	0.059	0.053

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A4: Spezifikation M.3 – Test auf unterdrückte Variablen

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	12	0.197	0.226	0.201	0.207	0.204	0.207	0.234	0.202
200	0	0.286	0.309	0.289	0.298	0.295	0.298	0.322	0.289
250	0	0.315	0.338	0.318	0.330	0.325	0.329	0.349	0.319
300	0	0.373	0.383	0.376	0.381	0.378	0.382	0.398	0.377
400	0	0.442	0.456	0.444	0.452	0.448	0.452	0.465	0.444
500	0	0.553	0.566	0.555	0.558	0.555	0.558	0.568	0.556
750	0	0.733	0.738	0.734	0.736	0.735	0.735	0.741	0.734
1500	0	0.947	0.947	0.947	0.947	0.947	0.947	0.948	0.947

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A5: Spezifikation M.3 – Test auf Heteroskedastie

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	12	0.047	0.075	0.120	0.078	0.065	0.069	0.116	0.061
200	0	0.049	0.068	0.091	0.068	0.062	0.067	0.103	0.058
250	0	0.049	0.069	0.085	0.067	0.063	0.067	0.095	0.059
300	0	0.049	0.067	0.079	0.067	0.064	0.066	0.087	0.058
400	0	0.054	0.065	0.074	0.065	0.063	0.065	0.080	0.059
500	0	0.050	0.061	0.067	0.063	0.058	0.059	0.075	0.055
750	0	0.050	0.057	0.064	0.059	0.059	0.059	0.071	0.054

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A6: Spezifikation M.2 – Test auf unterdrückte Variablen

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
150	9	0.059	0.080	0.067	0.079	0.077	0.076	0.092	0.069
200	0	0.056	0.068	0.057	0.060	0.062	0.060	0.071	0.057
250	0	0.054	0.062	0.055	0.058	0.061	0.058	0.067	0.055
300	0	0.052	0.062	0.054	0.056	0.056	0.056	0.067	0.054
400	0	0.055	0.060	0.056	0.058	0.060	0.057	0.061	0.056
500	0	0.052	0.055	0.053	0.055	0.058	0.055	0.059	0.053
750	0	0.052	0.057	0.052	0.053	0.054	0.053	0.059	0.052

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A7: Spezifikation M.4 – Test auf Heteroskedastie

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
200	2	0.052	0.064	0.100	0.077	0.071	0.073	0.129	0.062
250	0	0.051	0.064	0.088	0.071	0.064	0.068	0.111	0.061
300	0	0.052	0.063	0.081	0.068	0.065	0.064	0.113	0.060
400	0	0.048	0.057	0.068	0.060	0.054	0.057	0.102	0.054
500	0	0.050	0.056	0.064	0.058	0.056	0.057	0.097	0.054
750	0	0.050	0.052	0.059	0.056	0.055	0.054	0.100	0.053

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Tabelle A8: Spezifikation M.4 – Test auf unterdrückte Variablen

n	o.E.*	LM1	LM2	LM3	W1	W2	W3	W4	LR
200	2	0.052	0.060	0.054	0.056	0.056	0.056	0.064	0.054
250	0	0.047	0.051	0.047	0.050	0.049	0.050	0.055	0.047
300	0	0.050	0.055	0.051	0.054	0.055	0.053	0.058	0.051
400	0	0.051	0.053	0.052	0.052	0.053	0.053	0.055	0.052
500	0	0.047	0.048	0.047	0.048	0.048	0.048	0.049	0.047
750	0	0.053	0.054	0.054	0.055	0.054	0.055	0.056	0.054

* o.E. = Zahl der Replikationen bei denen der Schätzalgorithmus nicht konvergiert.

Abbildung 1a: Kerndichteschätzung, Quartic-Kern
LR-Test auf Heteroskedastie, Spezifikation M.1, $n = 250$

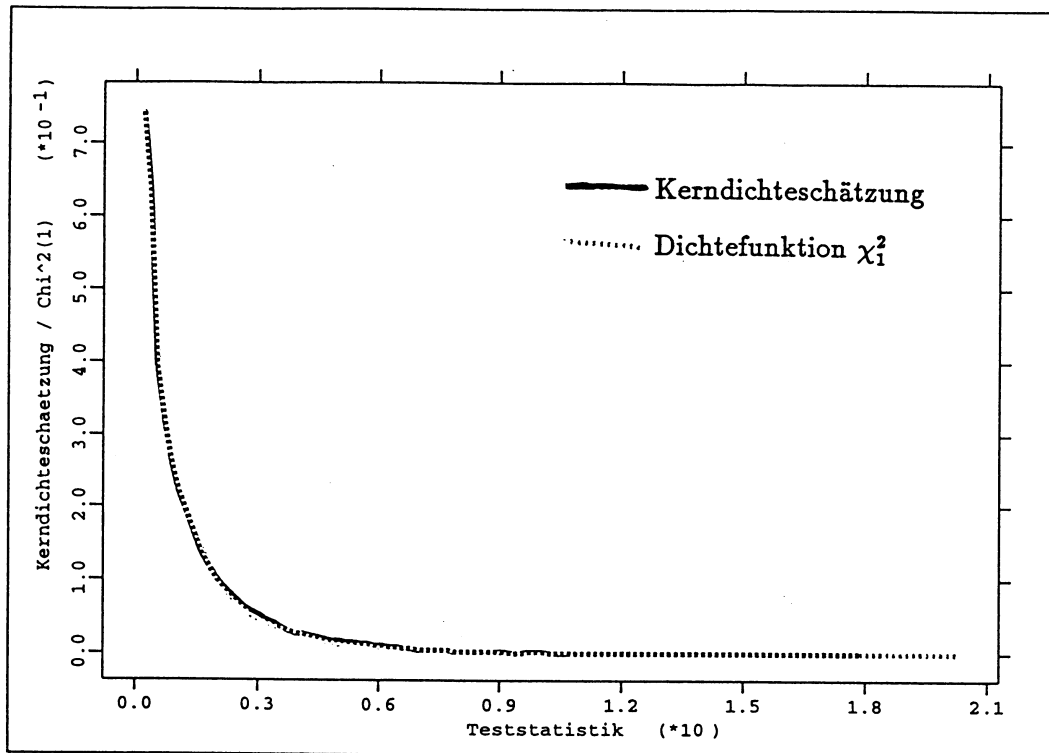


Abbildung 1b: Kerndichteschätzung, Quartic-Kern
LR-Test auf Heteroskedastie, Spezifikation M.1, $n = 750$

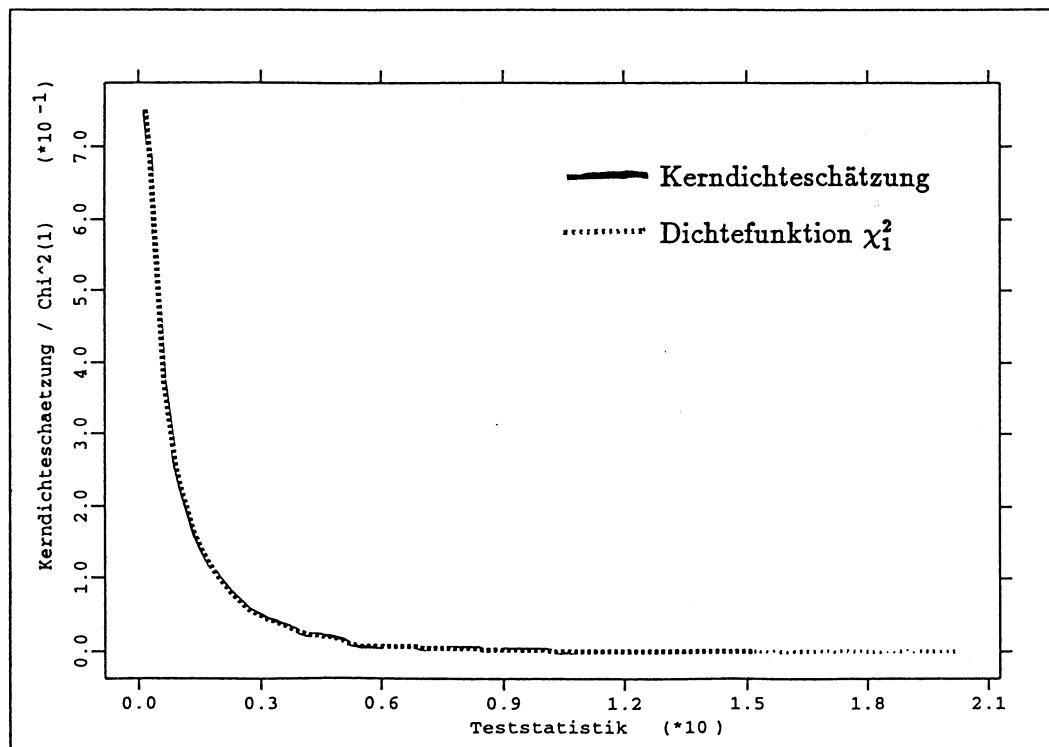


Abbildung 2:

LR-Test auf Heteroskedastie
Vergleich von Güte und Testniveau

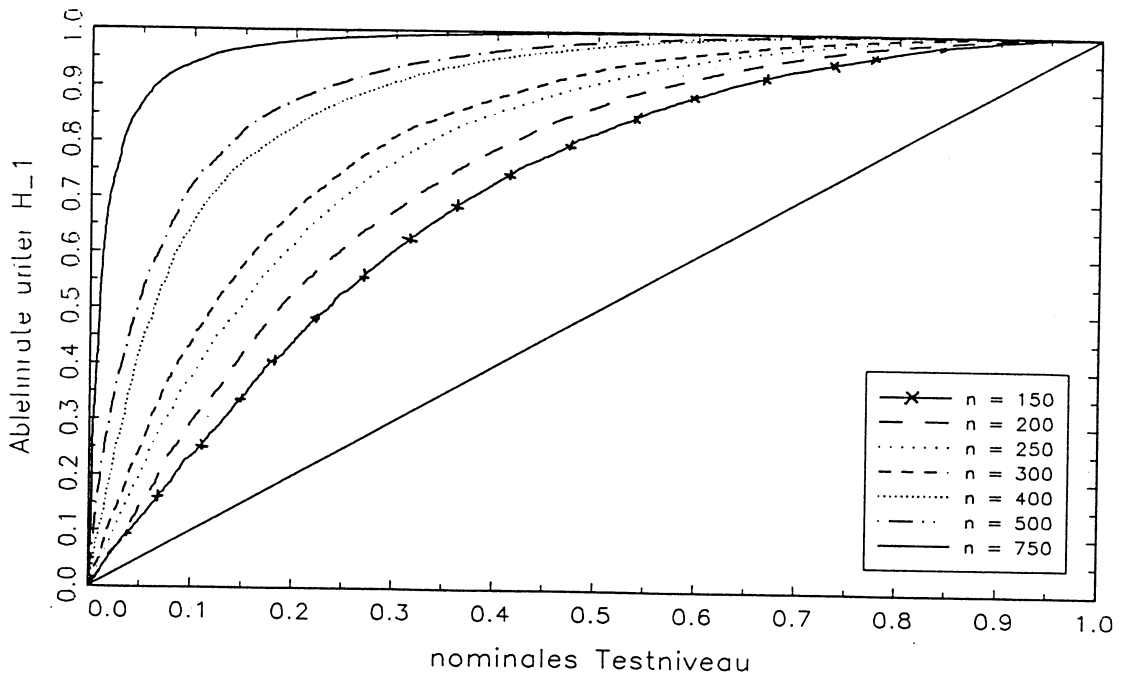


Abbildung 3:

LR-Test auf Heteroskedastie
Vergleich von Güte und Testniveau

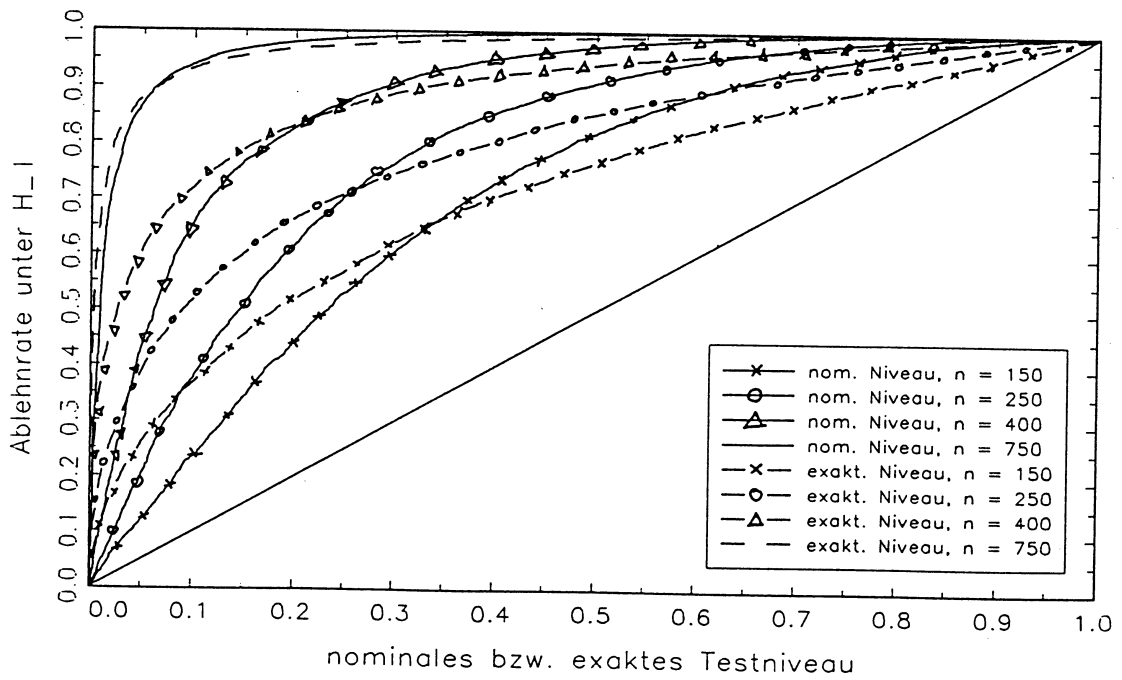
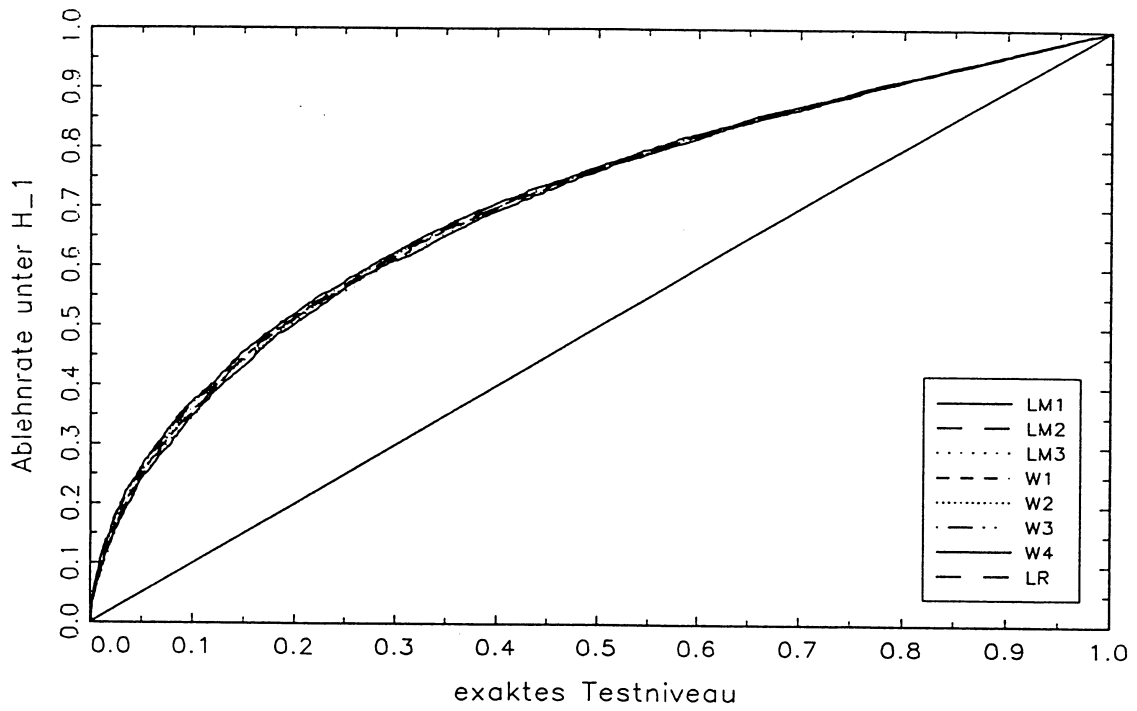


Abbildung 4:

Guetevergleich der Tests auf Heteroskedastie, $n = 150$



B Loglikelihoodfunktion eines Ordered Probit-Modells mit Heteroskedastie der Form $\sigma_i^2 = \sigma^2 \exp(2z_i'\lambda)$

Die Loglikelihoodfunktion lautet

$$\ln L = \sum_{i=1}^n \sum_{j=1}^J y_{ji} \ln(p_{ji}) \quad (\text{Anh.1})$$

mit

$$y_{ji} = \begin{cases} 1 & \text{wenn } y_i = j \\ 0 & \text{sonst} \end{cases}$$

$$p_{ji} := \Phi\left(\frac{\gamma_j - x_i'\beta}{\exp(z_i'\lambda)}\right) - \Phi\left(\frac{\gamma_{j-1} - x_i'\beta}{\exp(z_i'\lambda)}\right) . \quad (\text{Anh.2})$$

Dabei wurde in p_{ji} die Normierung $\sigma^2 = 1$ berücksichtigt.

B.1 Ableitungen 1. Ordnung

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} \frac{\partial p_{ji}}{\partial \beta}$$

$$\frac{\partial \ln L}{\partial \gamma_k} = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} \frac{\partial p_{ji}}{\partial \gamma_k}$$

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} \frac{\partial p_{ji}}{\partial \lambda}$$

mit

$$\frac{\partial p_{ji}}{\partial \beta} = -(\phi_{ji} - \phi_{(j-1)i}) \exp(-z_i'\lambda) x_i$$

$$\frac{\partial p_{ji}}{\partial \gamma_k} = (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k}) \exp(-z_i'\lambda)$$

$$\frac{\partial p_{ji}}{\partial \lambda} = -(\phi_{ji} g_{ji} - \phi_{(j-1)i} g_{(j-1)i}) z_i$$

und

$$\delta_{jk} = \begin{cases} 1 & \text{wenn } j = k \\ 0 & \text{sonst} \end{cases}$$

$$g_{ji} := \frac{\gamma_j - x'_i \beta}{\exp(z'_i \lambda)}$$

$$\phi_{ji} := \frac{1}{\exp(z'_i \lambda) \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\gamma_j - x'_i \beta}{\exp(z'_i \lambda)} \right)^2 \right]$$

$$= \phi \left(\frac{\gamma_j - x'_i \beta}{\exp(z'_i \lambda)} \right)$$

B.2 Ableitungen 2. Ordnung

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \beta \partial \beta'} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \beta} \frac{\partial p_{ji}}{\partial \beta'} \right)$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \gamma_k} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \beta \partial \gamma_k} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \beta} \frac{\partial p_{ji}}{\partial \gamma_k} \right) = \left(\frac{\partial^2 \ln L}{\partial \gamma_k \partial \beta'} \right)'$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \lambda'} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \beta \partial \lambda'} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \beta} \frac{\partial p_{ji}}{\partial \lambda'} \right) = \left(\frac{\partial^2 \ln L}{\partial \lambda \partial \beta'} \right)'$$

$$\frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_k} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \gamma_k \partial \gamma_k} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \gamma_k} \frac{\partial p_{ji}}{\partial \gamma_k} \right)$$

$$\frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_m} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \gamma_k \partial \gamma_m} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \gamma_k} \frac{\partial p_{ji}}{\partial \gamma_m} \right)$$

$$\frac{\partial^2 \ln L}{\partial \gamma_k \partial \lambda'} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \gamma_k \partial \lambda'} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \gamma_k} \frac{\partial p_{ji}}{\partial \lambda'} \right) = \left(\frac{\partial^2 \ln L}{\partial \lambda \partial \gamma_k} \right)'$$

$$\frac{\partial^2 \ln L}{\partial \lambda \partial \lambda'} = \sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \lambda \partial \lambda'} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \lambda} \frac{\partial p_{ji}}{\partial \lambda'} \right)$$

mit

$$\begin{aligned}
\frac{\partial^2 p_{ji}}{\partial \beta \partial \beta'} &= (\phi_{(j-1)i} g_{(j-1)i} - \phi_{ji} g_{ji}) \exp(-2z'_i \lambda) x_i x'_i \\
\frac{\partial^2 p_{ji}}{\partial \beta \partial \gamma_k} &= -(\phi_{(j-1)i} g_{(j-1)i} \delta_{(j-1)k} - \phi_{ji} g_{ji} \delta_{jk}) \exp(-2z'_i \lambda) x_i \\
\frac{\partial^2 p_{ji}}{\partial \beta \partial \lambda'} &= (\phi_{(j-1)i} (g_{(j-1)i}^2 - 2) - \phi_{ji} (g_{ji}^2 - 2)) \exp(-z'_i \lambda) x_i z'_i \\
\frac{\partial^2 p_{ji}}{\partial \gamma_k \partial \gamma_m} &= (\phi_{(j-1)i} g_{(j-1)i} \delta_{(j-1)k} \delta_{(j-1)m} - \phi_{ji} g_{ji} \delta_{jk} \delta_{jm}) \exp(-2z'_i \lambda) \\
&= 0 \quad \text{für } m \neq k \\
\frac{\partial^2 p_{ji}}{\partial \gamma_k \partial \lambda'} &= -(\phi_{(j-1)i} (g_{(j-1)i}^2 - 2) \delta_{(j-1)k} - \phi_{ji} (g_{ji}^2 - 2) \delta_{jk}) \exp(-z'_i \lambda) z'_i \\
\frac{\partial^2 p_{ji}}{\partial \lambda \partial \lambda'} &= (\phi_{(j-1)i} (g_{(j-1)i}^2 - 2) g_{(j-1)i} - \phi_{ji} (g_{ji}^2 - 2) g_{ji}) z_i z'_i
\end{aligned}$$

B.3 Informationsmatrix

Wegen

$$E(y_{ji}) = p_{ji} \Big|_{\theta=\theta_0}$$

und

$$\sum_{j=1}^J \frac{\partial^2 p_{ji}}{\partial \theta \partial \theta'} \Big|_{\theta=\theta_0} = 0$$

(vgl. z. B. Cramer 1986) gilt:

$$\begin{aligned}
-E \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right) \Big|_{\theta=\theta_0} &= -E \left(\sum_{i=1}^n \sum_{j=1}^J \left(\frac{y_{ji}}{p_{ji}} \frac{\partial^2 p_{ji}}{\partial \theta \partial \theta'} - \frac{y_{ji}}{p_{ji}^2} \frac{\partial p_{ji}}{\partial \theta} \frac{\partial p_{ji}}{\partial \theta'} \right) \right) \Big|_{\theta=\theta_0} \\
&= -E \left(\sum_{i=1}^n \sum_{j=1}^J \left(\frac{\partial^2 p_{ji}}{\partial \theta \partial \theta'} - \frac{1}{p_{ji}} \frac{\partial p_{ji}}{\partial \theta} \frac{\partial p_{ji}}{\partial \theta'} \right) \right) \Big|_{\theta=\theta_0} \\
&= E \left(\sum_{i=1}^n \sum_{j=1}^J \left(\frac{1}{p_{ji}} \frac{\partial p_{ji}}{\partial \theta} \frac{\partial p_{ji}}{\partial \theta'} \right) \right) \Big|_{\theta=\theta_0}
\end{aligned}$$

$$= E \left(\frac{\partial \ln L}{\partial \theta} \frac{\partial \ln L}{\partial \theta'} \right) \Big|_{\theta=\theta_0} .$$

Für die Elemente der Informationsmatrix $E \left(-\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right) \Big|_{\theta=\theta_0}$ folgt damit:

$$\begin{aligned} E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \beta'} \right) &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})^2}{p_{ji}} \exp(-2z'_i \lambda) x_i x'_i \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \gamma_k} \right) &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})(\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})}{p_{ji}} \\ &\quad \times \exp(-2z'_i \lambda) x_i \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \lambda'} \right) &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})(\phi_{ji} g_{ji} - \phi_{(j-1)i} g_{(j-1)i})}{p_{ji}} \\ &\quad \times \exp(-z'_i \lambda) x_i z'_i \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \gamma_k} \right) &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})^2}{p_{ji}} \exp(-2z'_i \lambda) \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \gamma_m} \right) &= 0 \quad \text{für } k \neq m \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \lambda'} \right) &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})}{p_{ji}} \\ &\quad \times (\phi_{ji} g_{ji} - \phi_{(j-1)i} g_{(j-1)i}) \exp(-z'_i \lambda) z'_i \\ E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \lambda} \frac{\partial \ln L_i}{\partial \lambda} \right) &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} g_{ji} - \phi_{(j-1)i} g_{(j-1)i})^2}{p_{ji}} z_i z'_i \end{aligned}$$

mit

$$\ln L_i := \sum_{j=1}^J y_{ji} \ln(p_{ji}) .$$

B.4 Ableitungen des Modells an der Stelle $\lambda = 0$

$$\frac{\partial \ln L}{\partial \beta} \Big|_{\lambda=0} = - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 - \phi_{(j-1)i}^0) x_i$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \gamma_k} \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \\ \frac{\partial \ln L}{\partial \lambda} \Big|_{\lambda=0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0) z_i\end{aligned}$$

mit

$$\begin{aligned}\phi_{ji}^0 &:= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\gamma_j - x'_i \beta)^2 \right] \\ &= \phi(\gamma_j - x'_i \beta) \\ g_{ji}^0 &:= (\gamma_j - x'_i \beta) \\ p_{ji}^0 &:= \Phi(\gamma_j - x'_i \beta) - \Phi(\gamma_{j-1} - x'_i \beta)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} x_i x'_i \\ \frac{\partial^2 \ln L}{\partial \beta \partial \gamma_k} \Big|_{\lambda=0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\} x_i \\ \frac{\partial^2 \ln L}{\partial \beta \partial \lambda'} \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 ([g_{(j-1)i}^0]^2 - 2) - \phi_{ji}^0 ([g_{ji}^0]^2 - 2)) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0) \right\} x_i z'_i \\ \frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_k} \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right.\end{aligned}$$

$$\begin{aligned}
& - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2 \Big\} \\
\left. \frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_m} \right|_{\lambda=0} &= 0 \quad \text{für } k \neq m \\
\left. \frac{\partial^2 \ln L}{\partial \gamma_k \partial \lambda'} \right|_{\lambda=0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 ([g_{(j-1)i}^0]^2 - 2) \delta_{(j-1)k} - \phi_{ji}^0 ([g_{ji}^0]^2 - 2) \delta_{jk}) \right. \\
& \quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0) \right\} z'_i \\
\left. \frac{\partial^2 \ln L}{\partial \lambda \partial \lambda'} \right|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 ([g_{(j-1)i}^0]^2 - 2) g_{(j-1)i}^0 - \phi_{ji}^0 ([g_{ji}^0]^2 - 2) g_{ji}^0) \right. \\
& \quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0)^2 \right\} z_i z'_i
\end{aligned}$$

Informationsmatrix:

$$\begin{aligned}
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \beta'} \right) \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} x_i x'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \gamma_k} \right) \Big|_{\lambda=0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})}{p_{ji}^0} x_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \lambda'} \right) \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0)}{p_{ji}^0} x_i z'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \gamma_k} \right) \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2}{p_{ji}^0} \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \lambda'} \right) \Big|_{\lambda=0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) (\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0)}{p_{ji}^0} z'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \lambda} \frac{\partial \ln L_i}{\partial \lambda} \right) \Big|_{\lambda=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 g_{ji}^0 - \phi_{(j-1)i}^0 g_{(j-1)i}^0)^2}{p_{ji}^0} z_i z'_i
\end{aligned}$$

mit

$$\ln L_i = \sum_{j=1}^J y_{ji} \ln(p_{ji}^0)$$

B.5 Loglikelihoodfunktion und Ableitungen unter der Nullhypothese $H_0 : \lambda = 0$

$$\ln L_0 = \sum_{i=1}^n \sum_{j=1}^J y_{ji} (\Phi(\gamma_j - x'_i \beta) - \Phi(\gamma_{j-1} - x'_i \beta)) \quad (\text{Anh.3})$$

$$\begin{aligned} \left. \frac{\partial \ln L_0}{\partial \beta} \right|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 - \phi_{(j-1)i}^0) x_i \\ \left. \frac{\partial \ln L_0}{\partial \gamma_k} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \\ \left. \frac{\partial^2 \ln L_0}{\partial \beta \partial \beta'} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} x_i x'_i \\ \left. \frac{\partial^2 \ln L_0}{\partial \beta \partial \gamma_k} \right|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} - (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\} x_i \\ \left. \frac{\partial^2 \ln L_0}{\partial \gamma_k \partial \gamma_k} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2 \right\} \\ \left. \frac{\partial^2 \ln L_0}{\partial \gamma_k \partial \gamma_m} \right|_{H_0} &= 0 \quad \text{für } k \neq m \end{aligned}$$

mit

$$\begin{aligned} \phi_{ji}^0 &:= \phi(\gamma_j - x'_i \beta) \\ g_{ji}^0 &:= (\gamma_j - x'_i \beta) \\ p_{ji}^0 &:= \Phi(\gamma_j - x'_i \beta) - \Phi(\gamma_{j-1} - x'_i \beta) \end{aligned}$$

Informationsmatrix:

$$\begin{aligned}
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \beta} \frac{\partial \ln L_{0,i}}{\partial \beta'} \right) \Big|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} x_i x_i' \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \beta} \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \right) \Big|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})}{p_{ji}^0} x_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \right) \Big|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2}{p_{ji}^0} \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \frac{\partial \ln L_{0,i}}{\partial \gamma_m} \right) \Big|_{H_0} &= 0 \quad \text{für } k \neq m
\end{aligned}$$

C Loglikelihoodfunktion eines Ordered Probit-Modells mit homoskedastischen Störtermen sowie den Re- gressoren x_i und w_i

Die Loglikelihoodfunktion lautet

$$\ln L = \sum_{i=1}^n \sum_{j=1}^J y_{ji} \ln(p_{ji}) \quad (\text{Anh.4})$$

mit

$$y_{ji} = \begin{cases} 1 & \text{wenn } y_i = j \\ 0 & \text{sonst} \end{cases}$$

$$p_{ji} := \Phi(\gamma_j - x_i' \beta - w_i' \alpha) - \Phi(\gamma_{j-1} - x_i' \beta - w_i' \alpha) \quad . \quad (\text{Anh.5})$$

Dabei wurde in p_{ji} die Normierung $\sigma^2 = 1$ berücksichtigt.

C.1 Ableitungen 1. Ordnung

$$\frac{\partial \ln L}{\partial \beta} = - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} (\phi_{ji} - \phi_{(j-1)i}) x_i$$

$$\frac{\partial \ln L}{\partial \gamma_k} = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})$$

$$\frac{\partial \ln L}{\partial \alpha} = - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}} (\phi_{ji} - \phi_{(j-1)i}) w_i$$

mit

$$\delta_{jk} = \begin{cases} 1 & \text{wenn } j = k \\ 0 & \text{sonst} \end{cases}$$

$$\begin{aligned}
g_{ji} &:= \gamma_j - x'_i \beta - w'_i \alpha \\
\phi_{ji} &:= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\gamma_j - x'_i \beta - w'_i \alpha)^2 \right] \\
&= \phi(\gamma_j - x'_i \beta - w'_i \alpha)
\end{aligned}$$

C.2 Ableitungen 2. Ordnung

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} - \phi_{ji} g_{ji}) \right. \\
&\quad \left. - \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} - \phi_{(j-1)i})^2 \right\} x_i x'_i \\
\frac{\partial^2 \ln L}{\partial \beta \partial \gamma_k} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} \delta_{(j-1)k} - \phi_{ji} g_{ji} \delta_{jk}) \right. \\
&\quad \left. - \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} - \phi_{(j-1)i}) (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k}) \right\} x_i \\
\frac{\partial^2 \ln L}{\partial \beta \partial \alpha'} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} - \phi_{ji} g_{ji}) \right. \\
&\quad \left. - \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} - \phi_{(j-1)i})^2 \right\} x_i w'_i \\
\frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_k} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} \delta_{(j-1)k} - \phi_{ji} g_{ji} \delta_{jk}) \right. \\
&\quad \left. - \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})^2 \right\} \\
\frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_m} &= 0 \quad \text{für } k \neq m \\
\frac{\partial^2 \ln L}{\partial \gamma_k \partial \alpha'} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} \delta_{(j-1)k} - \phi_{ji} g_{ji} \delta_{jk}) \right.
\end{aligned}$$

$$\frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} = \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}} (\phi_{(j-1)i} g_{(j-1)i} - \phi_{ji} g_{ji}) - \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} - \phi_{(j-1)i}) (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k}) \right\} w'_i$$

$$- \frac{y_{ji}}{p_{ji}^2} (\phi_{ji} - \phi_{(j-1)i})^2 \left. \right\} w_i w'_i$$

C.3 Informationsmatrix

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \beta'} \right) = \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})^2}{p_{ji}} x_i x'_i$$

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \gamma_k} \right) = - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i}) (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})}{p_{ji}} x_i$$

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \alpha'} \right) = \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})^2}{p_{ji}} x_i w'_i$$

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \gamma_k} \right) = \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})^2}{p_{ji}}$$

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \alpha'} \right) = - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i}) (\phi_{ji} \delta_{jk} - \phi_{(j-1)i} \delta_{(j-1)k})}{p_{ji}} w'_i$$

$$E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \alpha} \frac{\partial \ln L_i}{\partial \alpha'} \right) = \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji} - \phi_{(j-1)i})^2}{p_{ji}} w_i w'_i$$

mit

$$\ln L_i := \sum_{j=1}^J y_{ji} \ln(p_{ji}) \quad .$$

C.4 Ableitungen des Modells an der Stelle $\alpha = 0$

$$\left. \frac{\partial \ln L}{\partial \beta} \right|_{\alpha=0} = - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 - \phi_{(j-1)i}^0) x_i$$

$$\begin{aligned}\frac{\partial \ln L}{\partial \gamma_k} \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \\ \frac{\partial \ln L}{\partial \alpha} \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 - \phi_{(j-1)i}^0) w_i\end{aligned}$$

mit

$$\begin{aligned}\phi_{ji}^0 &:= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\gamma_j - x'_i \beta)^2 \right] \\ &= \phi(\gamma_j - x'_i \beta) \\ g_{ji}^0 &:= \gamma_j - x'_i \beta \\ p_{ji}^0 &:= \Phi(\gamma_j - x'_i \beta) - \Phi(\gamma_{j-1} - x'_i \beta)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} x_i x'_i \\ \frac{\partial^2 \ln L}{\partial \beta \partial \gamma_k} \Big|_{\alpha=0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\} x_i \\ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha'} \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} x_i w'_i \\ \frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_k} \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\}\end{aligned}$$

$$\begin{aligned}
\left. \frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_m} \right|_{\alpha=0} &= 0 && \text{für } k \neq m \\
\left. \frac{\partial^2 \ln L}{\partial \gamma_k \partial \alpha'} \right|_{\alpha=0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\
&\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\} w'_i \\
\left. \frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} \right|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\
&\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} w_i w'_i
\end{aligned}$$

Informationsmatrix:

$$\begin{aligned}
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \beta'} \right) \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} x_i x'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \gamma_k} \right) \Big|_{\alpha=0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})}{p_{ji}^0} x_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \beta} \frac{\partial \ln L_i}{\partial \alpha'} \right) \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} x_i w'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \gamma_k} \right) \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2}{p_{ji}^0} \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \gamma_k} \frac{\partial \ln L_i}{\partial \alpha'} \right) \Big|_{\alpha=0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})}{p_{ji}^0} w'_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_i}{\partial \alpha} \frac{\partial \ln L_i}{\partial \alpha'} \right) \Big|_{\alpha=0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} w_i w'_i
\end{aligned}$$

mit

$$\ln L_i := \sum_{j=1}^J y_{ji} \ln(p_{ji}^0) \quad .$$

C.5 Ableitungen unter $H_0 : \alpha = 0$

$$\ln L_0 = \sum_{i=1}^n \sum_{j=1}^J y_{ji} \ln (\Phi(\gamma_j - x'_i \beta) - \Phi(\gamma_{j-1} - x'_i \beta)) \quad (\text{Anh.6})$$

$$\begin{aligned} \left. \frac{\partial \ln L_0}{\partial \beta} \right|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 - \phi_{(j-1)i}^0) x_i \\ \left. \frac{\partial \ln L_0}{\partial \gamma_k} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ji}}{p_{ji}^0} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \\ \left. \frac{\partial^2 \ln L_0}{\partial \beta \partial \beta'} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 - \phi_{ji}^0 g_{ji}^0) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0)^2 \right\} x_i x'_i \\ \left. \frac{\partial^2 \ln L_0}{\partial \beta \partial \gamma_k} \right|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} - (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 - \phi_{(j-1)i}^0) (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k}) \right\} x_i \\ \left. \frac{\partial^2 \ln L_0}{\partial \gamma_k \partial \gamma_k} \right|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{y_{ji}}{p_{ji}^0} (\phi_{(j-1)i}^0 g_{(j-1)i}^0 \delta_{(j-1)k} - \phi_{ji}^0 g_{ji}^0 \delta_{jk}) \right. \\ &\quad \left. - \frac{y_{ji}}{(p_{ji}^0)^2} (\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2 \right\} \\ \left. \frac{\partial^2 \ln L}{\partial \gamma_k \partial \gamma_m} \right|_{H_0} &= 0 \quad \text{für } k \neq m \end{aligned}$$

mit

$$\begin{aligned} \phi_{ji}^0 &:= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\gamma_j - x'_i \beta)^2 \right] \\ &= \phi(\gamma_j - x'_i \beta) \\ g_{ji}^0 &:= (\gamma_j - x'_i \beta) \end{aligned}$$

$$p_{ji}^0 := \Phi(\gamma_j - x_i' \beta) - \Phi(\gamma_{j-1} - x_i' \beta)$$

Informationsmatrix

$$\begin{aligned}
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \beta} \frac{\partial \ln L_{0,i}}{\partial \beta'} \right) \Bigg|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)^2}{p_{ji}^0} x_i x_i' \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \beta} \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \right) \Bigg|_{H_0} &= - \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 - \phi_{(j-1)i}^0)(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})}{p_{ji}^0} x_i \\
E \left(\sum_{i=1}^n \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \frac{\partial \ln L_{0,i}}{\partial \gamma_k} \right) \Bigg|_{H_0} &= \sum_{i=1}^n \sum_{j=1}^J \frac{(\phi_{ji}^0 \delta_{jk} - \phi_{(j-1)i}^0 \delta_{(j-1)k})^2}{p_{ji}^0}
\end{aligned}$$

mit

$$\ln L_{0,i} = \sum_{j=1}^J y_{ij} \ln(p_{ij}^0)$$

D Lagrange-Multiplikator-Test

Sei $\hat{\theta}_{H_0}$ der unter einer Nullhypothese $H_0 : \theta_0 = \theta_{H_0}$ bestimmte Maximum-Likelihood-Schätzer für den unbekanntem Parametervektor θ_0 des Ordered Probit-Modells. Dann lautet die LM-Teststatistik auf diese Nullhypothese

$$LM = \frac{\partial \ln L_1}{\partial \theta} \Big|_{\theta=\hat{\theta}_{H_0}}' E \left(-\frac{\partial^2 \ln L_1}{\partial \theta \partial \theta'} \right)^{-1} \Big|_{\theta=\theta_{H_0}} \frac{\partial \ln L_1}{\partial \theta} \Big|_{\theta=\hat{\theta}_{H_0}}.$$

Damit ergibt sich für die Nullhypothese $H_0 : \lambda = 0$ mit dem ML-Schätzer $\hat{\theta}_{H_0} = (\hat{\beta}', \hat{\gamma}', 0)'$ für den Parametervektor $\theta_0 = (\beta', \gamma', \lambda)'$ und in Verbindung mit den Gradientenbedingungen für die Maximierung der Likelihoodfunktion

$$\begin{aligned} \frac{\partial \ln L_1}{\partial \beta} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \lambda=0} &= \frac{\partial \ln L_0}{\partial \beta} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \lambda=0} = 0 \\ \frac{\partial \ln L_1}{\partial \gamma_k} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \lambda=0} &= \frac{\partial \ln L_0}{\partial \gamma_k} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \lambda=0} = 0 \quad \forall k, \end{aligned}$$

wobei $\ln L_0$ die Loglikelihoodfunktion unter der Nullhypothese (Anh.3) und $\ln L_1$ die unter der Gegenhypothese (vgl. (Anh.1) und (Anh.2)) bezeichnet, die LM-Teststatistik

$$LM = \frac{\partial \ln L_1}{\partial \lambda} \Big|_{\theta=\hat{\theta}_{H_0}}' \cdot \Omega^{-1} \Big|_{\theta=\theta_{H_0}} \cdot \frac{\partial \ln L_1}{\partial \lambda} \Big|_{\theta=\hat{\theta}_{H_0}}$$

mit

$$\begin{aligned} \Omega^{-1} &:= -E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \lambda'} \right)^{-1} - E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \lambda'} \right)^{-1} \left[E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \beta'} \right) E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \gamma'} \right) \right] \\ &\times, \times \begin{bmatrix} E \left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \lambda'} \right) \\ E \left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \lambda'} \right) \end{bmatrix} \cdot E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \lambda'} \right)^{-1} \end{aligned} \quad (\text{Anh.7})$$

und

$$\begin{aligned}
, \quad & := \begin{bmatrix} E \left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \beta'} \right) & E \left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \gamma'} \right) \\ E \left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \beta'} \right) & E \left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \gamma'} \right) \end{bmatrix} \\
& - \begin{bmatrix} E \left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \lambda'} \right) \\ E \left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \lambda'} \right) \end{bmatrix} \cdot E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \lambda'} \right)^{-1} \begin{bmatrix} E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \beta'} \right) & E \left(\frac{\partial^2 \ln L_1}{\partial \lambda \partial \gamma'} \right) \end{bmatrix}
\end{aligned}$$

Die partiellen Ableitungen erster und zweiter Ordnung der Loglikelihoodfunktion unter der Null- und der Gegenhypothese sind wie in den Abschnitten B.4 und B.5 zu bilden. Die Elemente in $\mathbf{\Omega}^{-1}$ werden geschätzt durch die entsprechenden Elemente des äußeren Gradientenprodukts, der Hesseschen Matrix oder der Informationsmatrix, jeweils berechnet für $\hat{\theta}_{H_0}$.

Entsprechend folgt für den LM-Test auf die Hypothese $H_0 : \alpha = 0$ für den Parametervektor $\theta_0 = (\beta', \gamma', \alpha)'$ unter H_0 der ML-Schätzer $\hat{\theta}_{H_0} = (\hat{\beta}', \hat{\gamma}', 0)'$, und in Verbindung mit den Gradientenbedingungen für die Maximierung der Likelihoodfunktion

$$\begin{aligned}
\frac{\partial \ln L_1}{\partial \beta} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \alpha=0} &= \frac{\partial \ln L_0}{\partial \beta} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \alpha=0} = 0 \\
\frac{\partial \ln L_1}{\partial \gamma_k} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \alpha=0} &= \frac{\partial \ln L_0}{\partial \gamma_k} \Big|_{\beta=\hat{\beta}, \gamma=\hat{\gamma}, \alpha=0} = 0 \quad \forall k,
\end{aligned}$$

wobei wieder $\ln L_0$ die Loglikelihoodfunktion unter der Nullhypothese (Anh.6) und $\ln L_1$ die Loglikelihoodfunktion unter der Gegenhypothese (vgl. (Anh.4) und (Anh.5)) bezeichnet, folgt

$$LM = \frac{\partial \ln L_1}{\partial \alpha} \Big|_{\theta=\hat{\theta}_{H_0}}' \cdot \tilde{\mathbf{\Omega}}^{-1} \Big|_{\theta=\theta_{H_0}} \cdot \frac{\partial \ln L_1}{\partial \alpha} \Big|_{\theta=\hat{\theta}_{H_0}}$$

mit

$$\begin{aligned}
\tilde{\Omega}^{-1} &:= -E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \alpha'}\right)^{-1} - E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \alpha'}\right)^{-1} \left[E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \beta'}\right) E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \gamma'}\right) \right] \\
&\quad \times, \tilde{\times} \begin{bmatrix} E\left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \alpha'}\right) \\ E\left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \alpha'}\right) \end{bmatrix} \cdot E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \alpha'}\right)^{-1} \quad (\text{Anh.8})
\end{aligned}$$

und

$$\begin{aligned}
\tilde{\cdot} &:= \begin{bmatrix} E\left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \beta'}\right) & E\left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \gamma'}\right) \\ E\left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \beta'}\right) & E\left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \gamma'}\right) \end{bmatrix} \\
&\quad - \begin{bmatrix} E\left(\frac{\partial^2 \ln L_1}{\partial \beta \partial \alpha'}\right) \\ E\left(\frac{\partial^2 \ln L_1}{\partial \gamma \partial \alpha'}\right) \end{bmatrix} \cdot E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \alpha'}\right)^{-1} \left[E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \beta'}\right) E\left(\frac{\partial^2 \ln L_1}{\partial \alpha \partial \gamma'}\right) \right]
\end{aligned}$$

Dabei sind die partiellen Ableitungen erster und zweiter Ordnung in den Abschnitten C.4 und C.5 bestimmt. Die Elemente in $\tilde{\Omega}^{-1}$ werden wiederum geschätzt durch die entsprechenden Elemente des äußeren Gradientenprodukts, der Hesseschen Matrix oder der Informationsmatrix, jeweils berechnet für $\hat{\theta}_{H_0}$.

E Wald-Test

Sei $\hat{\theta}_{H_1}$ der unter einer Gegenhypothese $H_1 : \theta_0 \in \Theta_{H_1}$ bestimmte Maximum-Likelihood-Schätzer für den unbekanntem Parametervektor θ_0 des Ordered-Probit-Modells und $\ln L_1$ die zugehörige Loglikelihoodfunktion sowie $H_0 : \theta_0 = \theta_{H_0}$ eine zu testende Nullhypothese, die auch geschrieben werden kann als $H_0 : R(\theta_0) = r$. Dann lautet die Wald-Teststatistik auf diese Nullhypothese

$$W = (\hat{\theta}_{H_1} - \theta_{H_0})' \left\{ \frac{\partial R}{\partial \theta} \left[E \left(-\frac{\partial^2 \ln L_1}{\partial \theta \partial \theta'} \right)^{-1} \Big|_{\theta=\theta_0} \right] \frac{\partial R}{\partial \theta'} \right\}^{-1} (\hat{\theta}_{H_1} - \theta_{H_0}) .$$

Damit ergibt sich für die Nullhypothese $H_0 : \lambda = 0$ mit dem unter der Gegenhypothese bestimmten ML-Schätzer $\hat{\theta}_{H_1} = (\hat{\beta}', \hat{\gamma}', \hat{\lambda}')'$ für den Parametervektor $\theta_0 = (\beta', \gamma', \lambda)'$ die Wald-Teststatistik

$$W = \hat{\lambda}' \left\{ \mathbf{\Omega}^{-1} \Big|_{\theta=\theta_0} \right\}^{-1} \hat{\lambda}$$

mit $\mathbf{\Omega}^{-1}$ wie in Beziehung (Anh.7) definiert.

Analog folgt für den Test auf die Hypothese $H_0 : \alpha = 0$ mit dem unter H_1 berechneten ML-Schätzer $\hat{\theta}_{H_1} = (\hat{\beta}, \hat{\gamma}, \hat{\alpha})$ für den unbekanntem Parametervektor $\theta_0 = (\beta', \gamma', \alpha)'$ des Modells die Wald-Teststatistik

$$W = \hat{\alpha}' \left\{ \tilde{\mathbf{\Omega}}^{-1} \Big|_{\theta=\theta_0} \right\}^{-1} \hat{\alpha} .$$

Dabei ist $\tilde{\mathbf{\Omega}}^{-1}$ wie in Gleichung (Anh.8) definiert.