

Zur Quantifizierung und Analyse der Nichtlinearität von Regressionsmodellen

Günther Sedlacek
Wirtschaftsuniversität Wien

Zusammenfassung: Bei nichtlinearen Regressionsmodellen treffen im endlichen Stichprobenfall eine Reihe von statistischen Aussagen nicht zu, die für das lineare Regressionsmodell gelten. So ist die (iterativ berechnete) Kleinste-Quadrate-Summen-Schätzfunktion (KQS-Schätzfunktion) im nichtlinearen Fall nicht effizient, wenn sie auch unter gewissen Voraussetzungen asymptotisch effizient ist. Für nichtlineare Regressionsmodelle werden daher statistische Aussagen gerne mit Hilfe der asymptotischen Theorie getroffen. Zeigt das Regressionsmodell aber ausgeprägt nicht-lineares Verhalten, so führen so erhaltene Aussagen zu falschen Ergebnissen. Nichtlinearitätsmaße wie die von Bates and Watts (1980) geben Aufschluß darüber, wie ausgeprägt die Nichtlinearität eines Regressionsmodells ist. Passende Parametertransformationen sowie die Verwendung eines geeigneten Versuchsplans sind Möglichkeiten, die Nichtlinearität zu reduzieren, sodaß mit Hilfe von linearen Approximationen getroffene Aussagen zu validen Ergebnissen führen. Eine Fallstudie illustriert die Auswirkungen der Wahl der Versuchspunkte auf die Nichtlinearität eines Modells; die Verwendung eines geeigneten Versuchsplans kann zu einer deutlichen Reduzierung der Nichtlinearität führen.

Abstract: In nonlinear regression statistical analysis based upon interpretation of the parameter estimates may be quite different from linear regression. An important point is that for finite samples the least squares estimator (LSE) is not unbiased, nor it is a minimum variance estimator; for nonlinear models, the LSE has these properties under some assumptions only asymptotically and many statistical conclusions are based upon this asymptotic theory. But there are a lot of nonlinear models where the asymptotic properties are poorly approximated for finite samples. Assessing the nonlinearity (as in Bates and Watts, 1980) can show us if statistical tests, the justification of which rests on the assumption of linearity, are valid. Better parameterizations and experimental design are good possibilities to reduce the non-negligible nonlinearity of certain models. A case study shows that experimental design can reduce the nonlinearity considerably.

Schlüsselwörter: Nichtlineare Regression, Nichtlinearitätsmaße, Versuchsplanung.

1 Einleitung

Für viele Datensätze liefern lineare Regressionsmodelle nicht die gewünschte Anpassung. Nichtlineare Ertrag-Dichte Modelle und Wachstumsmodelle finden in der Praxis (z.B. in der Landwirtschaft) häufig Anwendung.

Wir betrachten nichtlineare Regressionsmodelle

$$Y_i = f(x_i, \boldsymbol{\theta}) + U_i, \quad i = 1, \dots, n, \quad (1)$$

wobei $\boldsymbol{\theta}$ der p -Vektor der unbekannt Parameter und x_i die i -te Messung der ein-dimensionalen unabhängigen Variable ist. Für die stochastischen Fehlerterme U_i wird angenommen, daß sie unabhängig und identisch mit Erwartungswert Null und konstanter, unbekannter Varianz σ^2 verteilt sind. Nichtlinearität wird in dem Sinn verstanden, daß die Regressionsfunktion $f(x_i, \boldsymbol{\theta})$ in zumindest einer Parameterkomponente nichtlinear ist.

Im weiteren gehen wir davon aus, daß die Kleinste-Quadrate-Summen (KQS)-Schätzfunktion asymptotisch effizient und normalverteilt ist. Seber and Wild (1989) geben jene Bedingungen (die sogenannten nichtlinearen Regularitätsbedingungen, engl. *nonlinear regularity conditions*) an, unter denen diese Annahmen gelten. Eine Reihe von statistischen Aussagen basiert auf der asymptotischen Theorie und führt daher bei Regressionsmodellen mit ausgeprägt nichtlinearem Verhalten für endliche Stichproben zu nicht validen Ergebnissen. Auch die Verzerrung der prognostizierten abhängigen Variable $\hat{Y}_i (= f(x_i, \hat{\boldsymbol{\theta}}))$, wobei $\hat{\boldsymbol{\theta}}$ ein Schätzer von $\boldsymbol{\theta}$ ist) hängt von der Nichtlinearität ab (siehe Ratkowsky, 1983). Es ist daher für die praktische Verwendung nichtlinearer Regressionsmodelle von Interesse, das Ausmaß des nichtlinearen Verhaltens zu ermitteln. Beschreibt ein Modell einen gegebenen Datensatz gut, weist jedoch ein hohes Maß an Nichtlinearität auf, so ist es wünschenswert, die Nichtlinearität zu verringern, um für den endlichen Stichprobenfall valide Aussagen aus der asymptotischen Theorie zu erhalten. Adäquate Parametertransformationen und geeignete Versuchspläne können die Nichtlinearität reduzieren.

Abschnitt 2 beschreibt kurz die wichtigsten Eigenschaften des nichtlinearen Regressionsmodells (Geometrie, KQS-Schätzer, Konfidenzregionen). Abschnitt 3 diskutiert verschiedene Nichtlinearitätsmaße sowie die Versuchsplanung im nichtlinearen Regressionsmodell und deren Effekt auf die Nichtlinearitätsmaße. Im vierten Abschnitt werden anhand eines ausgesuchten Datensatzes Nichtlinearitätswerte eines Wachstumsmodells ermittelt. Anschließend werden die Auswirkungen von Parametertransformationen und Veränderungen des Versuchsplans auf die Nichtlinearitätswerte des Modells untersucht.

Zur Vereinfachung wird folgende Notation eingeführt, wobei $i = 1, \dots, n$ und $j = 1, \dots, p$:

$$\begin{aligned} f_i(\boldsymbol{\theta}) &= f(x_i, \boldsymbol{\theta}), \\ \boldsymbol{\eta}(\boldsymbol{\theta}) &= (f_1(\boldsymbol{\theta}), \dots, f_n(\boldsymbol{\theta}))', \\ \dot{\mathbf{F}}(\boldsymbol{\theta}) &= \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left[\left(\frac{\partial f_i(\boldsymbol{\theta})}{\partial \theta_j} \right) \right]. \end{aligned}$$

Weiters verwenden wir

$$\dot{\mathbf{F}} = \dot{\mathbf{F}}(\boldsymbol{\theta}^*) \quad \text{und} \quad \hat{\dot{\mathbf{F}}} = \dot{\mathbf{F}}(\hat{\boldsymbol{\theta}}),$$

wobei $\boldsymbol{\theta}^*$ der Vektor der wahren Parameter, und $\hat{\boldsymbol{\theta}}$ der KQS-Schätzer ist.

Die Menge $\{\boldsymbol{\eta}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, Θ eine kompakte Teilmenge von \mathbb{R}^p , wird im Folgenden als Hyperfläche der Erwartungswerte bzw. als Wirkungshyperfläche (engl. *expectation surface, solution locus*) bezeichnet.

2 Das nichtlineare Regressionsmodell

In diesem Abschnitt werden die Effekte der Nichtlinearität anhand der grafischen Illustration eines einfachen nichtlinearen Regressionsmodells gezeigt und insbesondere die Auswirkungen auf die Ermittlung des KQS-Schätzers $\hat{\boldsymbol{\theta}}$ und von Konfidenzregionen für $\boldsymbol{\theta}$ behandelt.

2.1 Grafische Illustration der Nichtlinearität

Zur Illustration betrachten wir das nichtlineare Regressionsmodell

$$Y_i = x_i^\theta + U_i. \quad (2)$$

Das Modell soll an die beiden Beobachtungen $(2, 2.5)$ und $(3, 10)$ angepaßt werden. Als KQS-Schätzer (siehe Abschnitt 2.2) ergibt sich $\hat{\theta} = 2.054$, und $\hat{\sigma}^2 = 2.933$. Die eindimensionale Wirkungsfläche $\boldsymbol{\eta}(\theta)$ (Kurve) ist bestimmt durch

$$\boldsymbol{\eta}(\theta) = \begin{pmatrix} 2^\theta \\ 3^\theta \end{pmatrix}.$$

Im kartesischen Koordinatensystem mit Abszisse $\eta_1(\theta) = 2^\theta$ und Ordinate $\eta_2(\theta) = 3^\theta$ können wir $\boldsymbol{\eta}(\theta)$ unter Variation von θ darstellen (siehe Abbildung 1(a)); $\boldsymbol{\eta}(\theta)$ ist eine Kurve und die Abstände der Punkte auf $\boldsymbol{\eta}(\theta)$, die dem gleichen Inkrement $\Delta\theta$ entsprechen (in Abbildung 1(a) ist $\Delta\theta = 0.25$), sind nicht gleich. Im linearen Modell hingegen ist die eindimensionale Funktion $\boldsymbol{\eta}(\theta)$ eine Gerade und Werte, die gleichen Inkrementen $\Delta\theta$ entsprechen, liegen auf der Geraden $\boldsymbol{\eta}(\theta)$ auch wieder gleich weit voneinander entfernt.

Die grafische Präsentation nichtlinearer Modelle muß sich auf die Fälle $n \leq 3$ und $p \leq 2$ beschränken. Die für $p = 1$ gemachten Aussagen über den Unterschied zwischen linearen und nichtlinearen Modellen gelten auch für höhere Dimensionen. Ein lineares Regressionsmodell hat eine lineare Wirkungshyperfläche, d.h. für $p = 1$ eine Gerade, für $p = 2$ eine Ebene und für $p \geq 3$ eine Hyperebene. Auch die sogenannten Parameterlinien auf der Wirkungshyperfläche, das sind Linien, in denen eine Parameterkomponente Laufparameter ist und alle anderen Komponenten des Parametervektors konstant gehalten werden, zeigen die Linearität: Sie sind im linearen Fall gerade, parallel und haben für gleiche Inkremente dieser Parameterkomponente gleichen Abstand. Im nichtlinearen Fall hingegen ist die Wirkungshyperfläche gekrümmt, und die Parameterlinien sind weder gerade, noch parallel und gleich weit voneinander entfernt. Bates and Watts (1988) geben für $n = 3$ und $p = 2$ sowohl für den linearen als auch für den nichtlinearen Fall ein grafisches Beispiel.

Ein Maß für die Krümmung von $\boldsymbol{\eta}(\boldsymbol{\theta})$ in der Nähe der wahren Parameter $\boldsymbol{\theta}^*$ wird nach Beale (1960) und nach Bates and Watts (1980) als Intrinsische Nichtlinearität (engl. *intrinsic nonlinearity*) bezeichnet. Die Intrinsische Nichtlinearität mißt demnach, wie stark sich die Wirkungshyperfläche in der Nähe von $\boldsymbol{\theta}^*$ krümmt. Dieser Ausdruck ist deshalb passend, da sich die Intrinsische Nichtlinearität durch Parametertransformationen nicht verändert; die Gestalt von $\boldsymbol{\eta}(\boldsymbol{\theta})$ bleibt bei einer

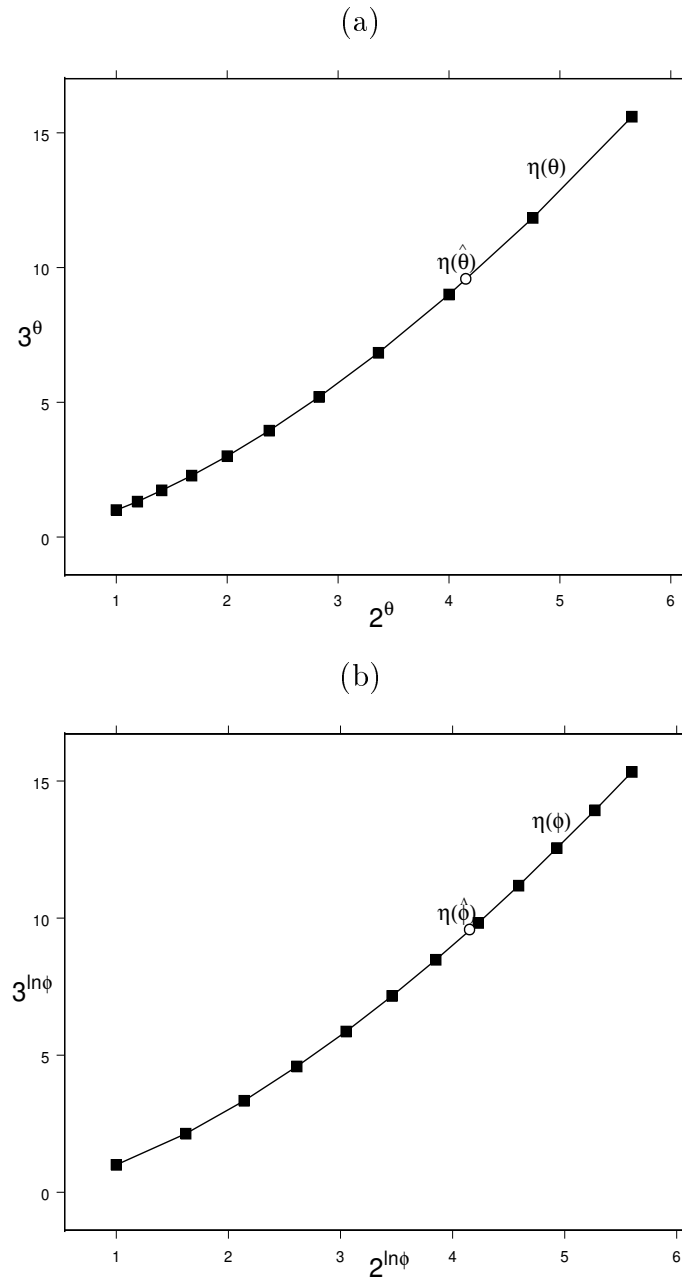


Abbildung 1: Eindimensionale Wirkungsfläche für $x_1 = 2$ und $x_2 = 3$ von (a) Modell (2) und (b) Modell (3). Die Punkte in (a) entsprechen den θ -Werten $\theta = 0, 0.25, 0.5, \dots, 2.5$ ($\Delta\theta = 0.25$) und in (b) den ϕ -Werten $\phi = 1, 2, \dots, 12$ ($\Delta\phi = 1$).

Parametertransformation unverändert. Im Falle von linearen Modellen oder nichtlinearen Modellen, die sich durch eine Parametertransformation linearisieren lassen, ist die Intrinsische Nichtlinearität gleich Null.

Ein Maß dafür, wie sehr gerade, parallele und gleich weit voneinander entfernte Parameterlinien in der Umgebung von $\boldsymbol{\theta}^*$ ihre Eigenschaften auf der Wirkungshyperfläche $\boldsymbol{\eta}(\boldsymbol{\theta})$ beibehalten, wird mit Parameter-Effekt-Nichtlinearität bezeichnet. Diese Bezeichnung ist deshalb passend, da die Parameter-Effekt-Nichtlinearität durch geeignete Parametertransformation stark reduziert werden kann. Ratkowsky (1983) zeigt anhand von zahlreichen Parametertransformationen zu unterschiedlichen nichtlinearen Regressionsmodellen, daß es häufig möglich ist, die Parameter-Effekt-Nichtlinearität deutlich zu verringern.

Um diesen Effekt zu demonstrieren, führen wir im Modell (2) einen neuen Parameter $\phi = \exp(\theta)$ ein und erhalten

$$Y_i = x_i^{\ln \phi} + U_i. \quad (3)$$

Abbildung 1(b) zeigt, daß die Punkte, die dem gleichen Inkrement $\Delta\phi = 1$ entsprechen, über den ganzen Bereich auf der Wirkungsfläche fast gleiche Abstände haben. Somit ist für Modell (3) die Parameter-Effekt-Nichtlinearität geringer als für Modell (2).

Verschiedene Regressionsfunktionen, die man durch unterschiedliche Parametrisierungen eines bestimmten Modells erhält, werden in der Literatur meist als Modellfunktionen bezeichnet (siehe Bates and Watts, 1980).

Im Abschnitt 3 wird eine Methode vorgestellt, die es ermöglicht, die Intrinsische und die Parameter-Effekt-Nichtlinearität eines Modells zu messen, und die aus den Grafiken erzielten Aussagen werden rechnerisch bestätigt.

2.2 Kleinste-Quadrate-Summen-Schätzer

Zur Ermittlung des KQS-Schätzers $\hat{\boldsymbol{\theta}}$ wird

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2$$

unter Variation von $\boldsymbol{\theta}$ minimiert, wobei y_i die i -te Beobachtung, $i = 1, \dots, n$, der eindimensionalen abhängigen Variable Y ist. Diese Aufgabe wird bei nichtlinearen Modellen im allgemeinen iterativ gelöst.

Die Standard-Methode ist die Gauss-Newton-Methode: Mit $\boldsymbol{\theta}^{(a)}$ wird die Approximation des KQS-Schätzers $\hat{\boldsymbol{\theta}}$ im nichtlinearen Modell bezeichnet, die sich in der a -ten Iteration ergibt. Auf ein beliebiges $\boldsymbol{\theta}$ nahe bei $\boldsymbol{\theta}^{(a)}$ wird eine lineare Taylorreihenentwicklung

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \approx \boldsymbol{\eta}(\boldsymbol{\theta}^{(a)}) + \dot{\mathbf{F}}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) \quad (4)$$

angewendet, wobei $\dot{\mathbf{F}}^{(a)} = \dot{\mathbf{F}}(\boldsymbol{\theta}^{(a)})$ ist. Analog gilt für den Fehlervektor $\mathbf{u}(\boldsymbol{\theta})$

$$\begin{aligned} \mathbf{u}(\boldsymbol{\theta}) &= \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}) \\ &\approx \mathbf{u}(\boldsymbol{\theta}^{(a)}) - \dot{\mathbf{F}}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}). \end{aligned}$$

Mit $S(\boldsymbol{\theta}) = \mathbf{u}'(\boldsymbol{\theta})\mathbf{u}(\boldsymbol{\theta})$ erhalten wir

$$S(\boldsymbol{\theta}) \approx \mathbf{u}'(\boldsymbol{\theta}^{(a)})\mathbf{u}(\boldsymbol{\theta}^{(a)}) - 2\mathbf{u}'(\boldsymbol{\theta}^{(a)})\dot{\mathbf{F}}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})'\dot{\mathbf{F}}^{(a)'}\dot{\mathbf{F}}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}). \quad (5)$$

Die rechte Seite wird minimiert in Bezug auf $\boldsymbol{\theta}$, falls gilt

$$\begin{aligned} \boldsymbol{\theta} - \boldsymbol{\theta}^{(a)} &= (\dot{\mathbf{F}}^{(a)'}\dot{\mathbf{F}}^{(a)})^{-1}\dot{\mathbf{F}}^{(a)'}\mathbf{u}(\boldsymbol{\theta}^{(a)}) \\ &=: \boldsymbol{\delta}^{(a)}. \end{aligned}$$

Somit erhält man die $(a + 1)$ -te iterative Verbesserung aus der a -ten Iteration zu

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + \boldsymbol{\delta}^{(a)}.$$

Der Gauss-Newton-Algorithmus ist Basis für viele Algorithmen zur iterativen Ermittlung des KQS-Schätzers. Es kann im Gauss-Newton-Algorithmus $S(\boldsymbol{\theta}^{(a+1)}) > S(\boldsymbol{\theta}^{(a)})$ eintreten, falls die lineare Approximation (4) nur in einer kleinen Umgebung von $\boldsymbol{\eta}(\boldsymbol{\theta}^{(a)})$ passend ist. In diesem Fall konvergiert $\boldsymbol{\theta}^{(a)}$ nicht gegen $\hat{\boldsymbol{\theta}}$. Eine geeignete Modifikation des Algorithmus wäre es z.B., eine Schrittweite λ einzuführen, sodaß

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + \lambda\boldsymbol{\delta}^{(a)}, \quad \text{und} \quad S(\boldsymbol{\theta}^{(a+1)}) < S(\boldsymbol{\theta}^{(a)}).$$

Im Unterschied zu linearen Modellen, bei denen die Minimierung der Summe der Fehlerquadrate in einem einzigen Schritt zu einem eindeutigen Minimum führt, ist es bei nichtlinearen Modellen möglich, daß $S(\boldsymbol{\theta})$ mehr als ein Minimum hat. Der verwendete Algorithmus kann divergieren oder zu einem lokalen Minimum - in der Nähe des Startwerts - und nicht zum globalen Minimum führen. Bei nichtlinearen Modellen, für die in jedem a -ten Schritt der Gauss-Newton-Methode die quadratische Approximation (5) eine gute Annäherung für $S(\boldsymbol{\theta})$ ist, tritt in jedem Fall - auch für Startwerte weit entfernt von $\hat{\boldsymbol{\theta}}$ - Konvergenz zum globalen Minimum in wenigen Schritten auf. Umso weniger ähnlich die Konturen von $S(\boldsymbol{\theta})$ und (5) sind, desto eher kann Divergenz bzw. Konvergenz zu einem lokalen Minimum in der Nähe des Startwerts auftreten. Geeignete Startwerte können entscheidend dazu beitragen, daß auch in diesen Fällen der Gauss-Newton Algorithmus bzw. Modifikationen davon zum globalen Minimum konvergieren (siehe Ratkowsky, 1983)).

In der Praxis bedeutet Konvergenz der iterativen Verfahren, daß sich $\boldsymbol{\theta}^{(a+1)}$ kaum mehr von $\boldsymbol{\theta}^{(a)}$ unterscheidet.

2.3 Konfidenzregionen

Mit $\hat{\boldsymbol{\beta}}$ als KQS-Schätzer des multiplen linearen Regressionsmodells, $\hat{\sigma}^2 = S(\hat{\boldsymbol{\beta}})/(n - p)$ und $F_{p, n-p; 1-\alpha}$ als $(1 - \alpha)$ -Quantile der $F(p, n - p)$ -Verteilung erhält man unter der Voraussetzung normalverteilter Fehlerterme als $100(1 - \alpha)\%$ Konfidenzregion für $\boldsymbol{\beta}$

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq p\hat{\sigma}^2 F_{p, n-p; 1-\alpha}.$$

Unter der Voraussetzung, daß die stochastischen Fehlerterme U_i identisch und unabhängig nach $N(0, \sigma^2)$ verteilt sind, und die nichtlinearen Regularitätsbedingungen

gelten, ist für $n \rightarrow \infty$ die Schätzfunktion $\hat{\boldsymbol{\theta}}$ asymptotisch nach $N_p(\boldsymbol{\theta}^*, \sigma^2(\dot{\mathbf{F}}'\dot{\mathbf{F}})^{-1})$ verteilt. Eine Schätzung der asymptotischen Varianz-Kovarianzmatrix $\sigma^2(\dot{\mathbf{F}}'\dot{\mathbf{F}})^{-1}$ erhalten wir, indem wir für $\dot{\mathbf{F}}$ die Matrix $\hat{\mathbf{F}} = \dot{\mathbf{F}}(\hat{\boldsymbol{\theta}})$ setzen. Die Matrix $\hat{\mathbf{F}}$ spielt dieselbe Rolle wie \mathbf{X} , die $(n \times k)$ -Matrix der unabhängigen Variablen, in der multiplen linearen Regression. Dementsprechend ergibt sich eine approximative $100(1 - \alpha)\%$ Konfidenzregion für $\boldsymbol{\theta}$ aus dem nichtlinearen Regressionsmodell (mit $\hat{\sigma}^2 = S(\hat{\boldsymbol{\theta}})/(n - p)$) zu

$$\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\hat{\mathbf{F}}'\hat{\mathbf{F}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq p\hat{\sigma}^2 F_{p, n-p; 1-\alpha}\}. \quad (6)$$

Eine alternative Möglichkeit, eine Konfidenzregion für $\boldsymbol{\theta}$ herzuleiten, kommt ohne Approximation der Regressionsfunktion aus und basiert auf der Likelihood-Funktion

$$p(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-1/2 \sum_{i=1}^n \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma^2}\right),$$

die von $\boldsymbol{\theta}$ nur über $S(\boldsymbol{\theta}) = \sum_i [y_i - f(x_i, \boldsymbol{\theta})]^2$ abhängt. Likelihood-Konturen sind für ein reelles $c > 1$ bestimmt durch $S(\boldsymbol{\theta}) = c$; sie bestehen somit aus allen Werten $\boldsymbol{\theta}$, für die $S(\boldsymbol{\theta})$ konstant ist. Likelihood-Konfidenzregionen sind daher von der Form

$$\{\boldsymbol{\theta} : S(\boldsymbol{\theta}) \leq cS(\hat{\boldsymbol{\theta}})\}.$$

Sie sind „exakt“, da sie nicht auf Approximationen basieren. Das Konfidenzniveau einer solchen Region jedoch kann nur asymptotisch mit Hilfe der Approximation

$$S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) \approx (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\hat{\mathbf{F}}'\hat{\mathbf{F}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

und (6) erhalten werden. Die $100(1 - \alpha)\%$ Likelihood-Konfidenzregion ergibt sich in Analogie zur linearen Theorie demnach näherungsweise zu

$$\left\{ \boldsymbol{\theta} : S(\boldsymbol{\theta}) \leq S(\hat{\boldsymbol{\theta}}) \left(1 + \frac{p}{n-p} F_{p, n-p; 1-\alpha} \right) \right\}. \quad (7)$$

Diese Definition gibt noch keinen Hinweis auf die Berechnung der Konfidenzregion, die im allgemeinen aufwendiger als die von Konfidenzregionen nach (6) ist.

Die Konfidenzregion (7) hat das geforderte asymptotische Signifikanzniveau von $100(1 - \alpha)\%$. Die Regionen (6) und (7) sind asymptotisch (wie auch im linearen Fall) die gleichen. Für den endlichen Stichprobenfall jedoch können sie verschieden sein. Während die Validität der linearisierten Konfidenzregionen (6) von der Intrinsischen und der Parameter-Effekt-Nichtlinearität abhängt, wird die Likelihood-Konfidenzregion (7) nur von der Intrinsischen Nichtlinearität beeinflusst, da $S(\boldsymbol{\theta})$ invariant gegenüber Parametertransformationen ist. Die linearisierten Konfidenzregionen (6) können auch für Modelle mit Intrinsischer Nichtlinearität gleich Null inakzeptabel sein; das Konfidenzniveau der Likelihood-Region hingegen ist in diesem Fall exakt $(1 - \alpha)$. Sind die beiden Regionen sehr unterschiedlich, so deutet dies darauf hin, daß die lineare (lokale) Approximation der Wirkungshyperfläche $\boldsymbol{\eta}(\boldsymbol{\theta})$ an $\hat{\boldsymbol{\theta}}$, also

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \approx \boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) + \dot{\mathbf{F}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \quad (8)$$

und somit (6) nicht geeignet sind. Für den Fall von kleiner Intrinsischer Nichtlinearität und ausgeprägter Parameter-Effekt-Nichtlinearität sind die Likelihood-Regionen weit passender als (6). Sind beide Nichtlinearitäten klein, empfiehlt es sich, wegen der leichteren Berechenbarkeit (6) zu verwenden. Alternative Konfidenzregionen, die auch für Modelle mit ausgeprägter Intrinsischer und Parameter-Effekt-Nichtlinearität valide Ergebnisse liefern, werden z.B. in Pázman (1993) oder in Seber and Wild (1989) behandelt. Der nächste Abschnitt diskutiert Methoden, wie die Nichtlinearität eines Modells quantitativ untersucht und reduziert werden kann, und gibt den Begriffen „klein“ und „ausgeprägt“ einen Wert.

In der Praxis läuft man natürlich Gefahr, den Resultaten aus der asymptotischen Theorie zu viel Vertrauen zu schenken. Schon bei den iterativen Verfahren zur Berechnung des KQS-Schätzers ist erwähnt worden, daß die Konvergenz der Methoden davon abhängt, wie passend die lineare Approximation (4) für jeden iterativen Schritt ist. Auch die Brauchbarkeit der Konfidenzregion (6) und das Verhalten der KQS-Schätzfunktion hängen davon ab, wie adäquat eine lineare Approximation von $\boldsymbol{\eta}(\boldsymbol{\theta})$ in der Nähe von $\hat{\boldsymbol{\theta}}$ ist. Unter den nichtlinearen Regularitätsbedingungen ist die KQS-Schätzfunktion wie eingangs erwähnt asymptotisch effizient und (multivariat) normalverteilt. Ist die lineare Approximation nicht passend, so wird die KQS-Schätzfunktion im endlichen Stichprobenfall stark verzerrt sein, eine erheblich abweichende Verteilung von der Normalverteilung und eine Varianz haben, die die minimale Varianzschranke weit übertrifft.

3 Nichtlinearitätsmaße

Nichtlinearitätsmaße sind zahlenmäßige Indikatoren, ob eine lineare Approximation und die auf ihr basierenden statistischen Aussagen für endliche Stichproben adäquat sind. Im Anschluß an die in Abschnitt 2.1 geführte Diskussion der Ideen, die solchen Nichtlinearitätsmaßen zugrunde liegen, werden in diesem Abschnitt die Nichtlinearitätsmaße von Bates and Watts (1980) und Box (1971) definiert. Andere Indikatoren für die Nichtlinearität eines Regressionsmodells finden sich bei Beale (1960) und Ratkowsky (1983). Schließlich werden die durch Veränderung des Versuchsplans erzielten Effekte auf die Nichtlinearität diskutiert.

3.1 Nichtlinearitätsmaße von Bates and Watts

Bates and Watts (1980) entwickelten ihre Nichtlinearitätsmaße aus Ideen der Differentialgeometrie.

Ein grundlegendes Charakteristikum linearer Modelle ist, daß die zweite und daher auch alle höheren Ableitungen von $f(x_i, \boldsymbol{\theta})$ nach den Parametern Null sind. Die Nichtlinearitätsmaße von Bates and Watts (1980) basieren auf den zweiten Ableitungen von $\boldsymbol{\eta}(\boldsymbol{\theta})$. Es wird folgende Notation verwendet

$$\dot{\mathbf{F}} = \left[\left(\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_r} \right) \right] = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

und

$$\ddot{\mathbf{F}} = \left[\left(\frac{\partial^2 f(x_i, \boldsymbol{\theta})}{\partial \theta_r \partial \theta_s} \right) \right] = \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

mit $r, s = 1, \dots, p$ und $i = 1, \dots, n$. $\dot{\mathbf{F}}$ besteht somit aus den n -Tangentenvektoren $\dot{\mathbf{f}}_r = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_r}$, $r = 1, \dots, p$; die i -te Scheibe (engl. *face*) $\dot{\mathbf{F}}_i$ von $\dot{\mathbf{F}}$, $i = 1, \dots, n$, ist eine $(p \times p)$ -Matrix; $\dot{\mathbf{F}}$ ist daher ein n -Tupel von $(p \times p)$ -Matrizen, das aus den Vektoren $\ddot{\mathbf{f}}_{rs} = \frac{\partial \dot{\mathbf{f}}_r}{\partial \theta_s}$, $r, s = 1, \dots, p$, besteht, d.h. eine $(n \times p \times p)$ -Matrix. Die Tangentenvektoren $\dot{\mathbf{f}}$ und die Vektoren $\ddot{\mathbf{f}}$ werden in der englischen Literatur meist als *velocity-* bzw. *acceleration-vectors* bezeichnet.

Eine beliebige Gerade im Parameterraum durch $\hat{\boldsymbol{\theta}}$ mit beliebigem $\delta \neq 0 \in \mathbb{R}$ hat folgende Gestalt

$$\boldsymbol{\theta}(\delta) = \hat{\boldsymbol{\theta}} + \delta \mathbf{h},$$

wobei $\mathbf{h} = (h_1, \dots, h_p)'$ die Richtung der Geraden angibt. Diese Gerade erzeugt auf der Wirkungshyperfläche folgende Kurve

$$\boldsymbol{\eta}_h(\boldsymbol{\theta}(\delta)) = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}} + \delta \mathbf{h})$$

durch $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$. Der Tangentenvektor für diese Kurve an $\hat{\boldsymbol{\theta}}$ ist

$$\begin{aligned} \dot{\boldsymbol{\eta}}_h &= \frac{d\boldsymbol{\eta}_h}{d\delta} \Big|_0 = \sum_{r=1}^p \frac{\partial \boldsymbol{\eta}}{\partial \theta_r} \Big|_{\hat{\boldsymbol{\theta}}} \frac{d\theta_r}{d\delta} \Big|_0 \\ &= \sum_{r=1}^p \dot{\mathbf{f}}_r h_r, \end{aligned}$$

oder $\dot{\boldsymbol{\eta}}_h = \dot{\mathbf{F}} \mathbf{h}$; $\dot{\boldsymbol{\eta}}_h$ ist eine Linearkombination von den Vektoren $\dot{\mathbf{f}}$. Somit ist der Bereich der Matrix $\dot{\mathbf{F}}$ die Tangentialhyperebene an $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$.

Ähnlich erhält man die zweite Richtungsableitung von $\boldsymbol{\eta}_h$ an $\delta = 0$

$$\begin{aligned} \ddot{\boldsymbol{\eta}}_h &= \frac{d^2 \boldsymbol{\eta}_h}{d\delta^2} \Big|_0 = \sum_{s=1}^p \frac{\partial \sum_{r=1}^p \dot{\mathbf{f}}_r h_r}{\partial \theta_s} \Big|_{\hat{\boldsymbol{\theta}}} \frac{d\theta_s}{d\delta} \Big|_0 \\ &= \sum_{r=1}^p \sum_{s=1}^p \ddot{\mathbf{f}}_{rs} h_r h_s, \end{aligned}$$

oder $\ddot{\boldsymbol{\eta}}_h = \mathbf{h}' \ddot{\mathbf{F}} \mathbf{h}$; $\ddot{\boldsymbol{\eta}}_h$ ist eine Linearkombination der Vektoren $\ddot{\mathbf{f}}$.

Der Vektor $\ddot{\boldsymbol{\eta}}_h$ kann in zwei orthogonale Komponenten $\ddot{\boldsymbol{\eta}}_h^N = \mathbf{h}' \ddot{\mathbf{F}}^N \mathbf{h}$ und $\ddot{\boldsymbol{\eta}}_h^T = \mathbf{h}' \ddot{\mathbf{F}}^T \mathbf{h}$ zerlegt werden, von denen $\ddot{\boldsymbol{\eta}}_h^N$ normal und $\ddot{\boldsymbol{\eta}}_h^T$ parallel zur Tangentialebene liegen.

Als (mit $\rho = \hat{\sigma} \sqrt{p}$) standardisierte relative Intrinsische Nichtlinearität bezüglich der Richtung \mathbf{h} definieren Bates and Watts (1980)

$$\gamma_h^N = \rho \frac{\|\ddot{\boldsymbol{\eta}}_h^N\|}{\|\dot{\boldsymbol{\eta}}_h\|^2}.$$

Die Intrinsische Nichtlinearität kann als der inverse Radius jenes Kreises, der die Wirkungshyperfläche in Richtung $\dot{\boldsymbol{\eta}}_h$ an $\hat{\boldsymbol{\theta}}$ am besten approximiert, interpretiert

werden; $\gamma_{\mathbf{h}}^N$ hängt von der Gestalt von $\boldsymbol{\eta}(\boldsymbol{\theta})$ und vom Versuchsplan, jedoch nicht von der verwendeten Parametrisierung ab.

Analog wird als standardisierte relative Parameter-Effekt-Nichtlinearität

$$\gamma_{\mathbf{h}}^T = \rho \frac{\|\ddot{\boldsymbol{\eta}}_{\mathbf{h}}^T\|}{\|\dot{\boldsymbol{\eta}}_{\mathbf{h}}\|^2}$$

definiert. Um jeweils einen einzigen, von der Richtung \mathbf{h} unabhängigen Wert für die Intrinsische und für die Parameter-Effekt-Nichtlinearität zu erhalten, schlagen Bates and Watts vor, jeweils die Maxima

$$\Gamma^N = \max_{\mathbf{h}} \gamma_{\mathbf{h}}^N \quad \text{und} \quad \Gamma^T = \max_{\mathbf{h}} \gamma_{\mathbf{h}}^T$$

zu ermitteln. Zur konkreten Berechnung wird für $\boldsymbol{\theta}$ ein Schätzer $\hat{\boldsymbol{\theta}}$ eingesetzt. Einen Algorithmus zur praktischen Berechnung liefern Bates and Watts (1980), und eine ausführliche geometrische Interpretation und Herleitung dieser Nichtlinearitätsmaße geben Bates and Watts (1988) und Pázman (1993).

Ein Nachteil dieser Nichtlinearitätsmaße ist, daß sie die maximale Nichtlinearität im Modell angeben. Eine Alternative ist die mittlere Quadrat-Nichtlinearität (engl. *root-mean-square curvature*)

$$(\gamma_{RMS}^N)^2 = \frac{1}{S(p)} \int_{\|\mathbf{h}\|=1} (\gamma_{\mathbf{h}}^N)^2 dS,$$

wobei $\mathbf{h} = (h_1, \dots, h_p)$ und $S(p)$ die Oberfläche der p -dimensionalen Einheitskugel ist. Sie mißt eine durchschnittliche Nichtlinearität bezüglich allen Richtungen.

Die lineare Approximation ist im endlichen Fall dann geeignet, falls Γ^N und Γ^T (bzw. γ_{RMS}^N und γ_{RMS}^T) klein sind. Wirklich aussagekräftig werden die beiden Krümmungsbegriffe aber erst, wenn ein passender Vergleichswert vorhanden ist, sodaß der Begriff „klein“ quantifizierbar wird. Dazu betrachten wir die Konfidenzregion (6)

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \hat{\mathbf{F}}' \hat{\mathbf{F}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq p \hat{\sigma}^2 F_{p, n-p; 1-\alpha}.$$

Das Skalieren mit $p \hat{\sigma}^2$ ergibt aus (6) eine Konfidenzregion mit Radius $\sqrt{F} := \sqrt{F_{p, n-p; 1-\alpha}}$; für die Krümmung (= inversen Radius) der Konfidenzregion erhält man daher $1/\sqrt{F}$. Bates and Watts schlagen aufgrund zahlreicher Studien vor, die Nichtlinearitäten als ausgeprägt anzusehen, falls $\Gamma^T > \Gamma_0$ bzw. $\Gamma^N > \Gamma_0$, wobei $\Gamma_0 = \frac{1}{2\sqrt{F}}$. Eine geometrische Interpretation des Vergleichswertes Γ_0 liefern Bates and Watts (1980) und Seber and Wild (1989). Ist die Intrinsische Nichtlinearität nicht ausgeprägt, kann die Parameter-Effekt-Nichtlinearität untersucht werden. Ist auch diese nicht ausgeprägt, so ist eine lineare Approximation der Wirkungshyperfläche adäquat und die Ergebnisse aus der linearen Approximation (wie z.B. (6)) plausibel. Ist die Intrinsische Nichtlinearität geringfügig und die Parameter-Effekt-Nichtlinearität hoch, so kann diese z.B. mit Hilfe von Parametertransformationen verringert werden. Die Nichtlinearitätsmaße geben aber keinen Hinweis, welche Parameterkomponenten am meisten zur Nichtlinearität beitragen. Seber and Wild

(1989) behandeln den direkten Effekt der Nichtlinearitätsmaße von Bates and Watts auf die Validität der linearisierten Konfidenzregionen (6).

Für Modell (2) und $x_1 = 2, x_2 = 3$ erhalten wir $\Gamma^T = 0.169$ und $\Gamma^N = 0.016$. Die Intrinsische Nichtlinearität ist zu vernachlässigen ($\Gamma_0 = 0.039 > \Gamma^N$), die Parameter-Effekt-Nichtlinearität hingegen ist ausgeprägt ($\Gamma^T > \Gamma_0$). Die Reparametrisierung (3) liefert $\Gamma^T = 0.011 < \Gamma_0$ und reduziert die Parameter-Effekt-Nichtlinearität deutlich.

3.2 Die Verzerrung nach M.J. Box

Die Nichtlinearität eines Modells kann auch am Verhalten der KQS-Schätzfunktion (für endliche Stichproben) gemessen werden. Es werden wieder identisch und unabhängig normal verteilte Fehlerterme U_i vorausgesetzt, und die Stichprobeneigenschaften von $\hat{\boldsymbol{\theta}}$ werden komponentenweise untersucht.

Bezeichnet man mit $\dot{\mathbf{F}}_i$ ($= \dot{\mathbf{F}}_u$) den Vektor der ersten Ableitungen von $f(x_i, \boldsymbol{\theta})$ nach $\boldsymbol{\theta}$, also die i -te Zeile, $i = 1, \dots, n$, von $\dot{\mathbf{F}}$, und mit $\ddot{\mathbf{F}}_i$ die i -te Matrixscheibe von $\ddot{\mathbf{F}}$, so erhält man nach Box (1971) für die Verzerrung ($B(\boldsymbol{\theta}) := E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$) der Parameter in einem nichtlinearen Modell

$$B(\boldsymbol{\theta}) = \frac{-\sigma^2}{2} \left(\sum_{u=1}^n \dot{\mathbf{F}}'_u \dot{\mathbf{F}}_u \right)^{-1} \sum_{i=1}^n \dot{\mathbf{F}}'_i \text{tr} \left[\left(\sum_{u=1}^n \dot{\mathbf{F}}'_u \dot{\mathbf{F}}_u \right)^{-1} \ddot{\mathbf{F}}_i \right]. \quad (9)$$

Für die Berechnung von (9) setzt man für die unbekannt Parameter die Schätzer $\hat{\boldsymbol{\theta}}$ und $\hat{\sigma}^2$ ein und erhält für die Verzerrung von $\hat{\boldsymbol{\theta}}$ einen p -Vektor $B(\boldsymbol{\theta})$. Bates and Watts (1980) zeigen, daß die Verzerrung nach Box in direkter Relation zu ihrem Maß für die Parameter-Effekt-Krümmung steht; $B(\boldsymbol{\theta})$ hängt nur von der Parametrisierung und nicht von der Intrinsischen Nichtlinearität ab und kann häufig durch eine geeignete Reparametrisierung der Regressionsfunktion reduziert werden.

In der Praxis werden die Komponenten des Vektors $B(\boldsymbol{\theta})$ meist nicht als Absolutwerte sondern als Prozentwerte

$$\%B(\theta_j) = \frac{B(\theta_j)}{\theta_j}, \quad j = 1, \dots, p,$$

angegeben. Folgende Regel hat sich als gute Faustregel erwiesen: Ist $\%B(\theta_j)$ für einen Parameter absolut größer als 1%, so ist dies ein Indikator für nichtlineares Verhalten des Modells in θ_j (siehe Ratkowsky, 1983).

Die Verzerrung $\%B(\theta_j)$ gibt somit einen Hinweis darauf, welche Beiträge die Parameterkomponenten zur Nichtlinearität eines Modells leisten. Sie gibt aber keinen Hinweis, wie Komponenten mit hohem $\%B$ transformiert werden sollen, um die Parameter-Effekt-Nichtlinearität zu verringern. Wird von einer alternativen Reparametrisierung, z.B. aus Simulationsstudien (Ratkowsky, 1983), angenommen, daß diese eine kleinere (nicht ausgeprägte) Parameter-Effekt-Nichtlinearität hervorbringt, so läßt sich die Verzerrung der „neuen“ Parameter direkt aus der der „alten“ Parameter bestimmen (siehe Box, 1971).

3.3 Effekte des Versuchsplans

Sowohl die Intrinsische als auch die Parameter-Effekt-Nichtlinearität sind von der Wahl der Versuchspunkte x_i abhängig, was sich am Beispiel der Veränderung des Versuchsplans zeigen läßt. Dazu betrachten wir zur Illustration Modell (2) für den neuen Versuchsplan $x_1 = 1.1$ und $x_2 = 1.3$. Die konkreten Berechnungen der Nichtlinearitätswerte basieren auf der Überlegung, daß für die wahren Parameter θ^* und σ^2 die für das Modell (2) bereits berechneten Parameterschätzer $\hat{\theta} = 2.054$ und $\hat{\sigma}^2 = 2.933$ gesetzt werden. Abbildung 2 zeigt für den neuen Versuchsplan die eindimensionale Wirkungsfläche für die Modelle (2) und (3). Die Nichtlinearitätswerte nach Bates and Watts ergeben sich für Modell (2) zu $\Gamma^T = 0.929$ und $\Gamma^N = 0.149$; beide Werte sind für den neuen Versuchsplan größer. Wie aus Abbildung 2(a) ersichtlich, führt die Reparametrisierung (3) für den neuen Versuchsplan nicht zu einer Verringerung sondern zu einer weiteren Erhöhung des Parameter-Effekt-Nichtlinearitätswertes ($\Gamma^T = 2.758$).

Es ist daher naheliegend, geeignete Versuchspläne zu suchen, die zu einer deutlichen Reduzierung der Nichtlinearität eines Modells führen. Ratkowsky (1983) zeigt an einem Beispiel, daß schon alleine die gleichmäßige Verteilung der Beobachtungen über den vermuteten Wertebereich eine beträchtliche Verringerung der Parameter-Effekt-Nichtlinearität bewirken kann. Viele Autoren setzen sich aber vorwiegend mit der Versuchsplanung zur möglichst präzisen Schätzung der unbekannt Parameter auseinander. Insbesondere das D-Optimalitätskriterium findet häufig seine Anwendung.

Ein Versuchsplan

$$X = \begin{pmatrix} x_1 & , & \dots & , & x_m \\ n_1 & , & \dots & , & n_m \end{pmatrix}$$

mit n_k ($\sum n_k = n$) Beobachtungen an der Stelle x_k , $k = 1, \dots, m$, heißt analog zu linearen Modellen (mit $\mathbf{X} = \mathbf{F}$) lokal D-optimal, falls er die Determinante der asymptotischen Varianz-Kovarianzmatrix von $\hat{\theta}$, also

$$\det(\dot{\mathbf{F}}'\dot{\mathbf{F}})^{-1}$$

minimiert. Für nichtlineare Modelle ist dieses Kriterium im Unterschied zu linearen Versuchsplänen von den wahren Parametern θ^* abhängig und daher nur für das ausgewählte θ (nahe θ^*) lokal optimal. Die große Akzeptanz von D-optimalen Versuchsplänen kommt daher, daß sie das Volumen der linearisierten Konfidenzregion (6) minimieren und invariant gegenüber der verwendeten Skalierung der Parameter sind.

Box and Lucas (1959) zeigen in ihrer Pionierarbeit zur lokal optimalen Versuchsplanung anhand eines einfachen nichtlinearen Regressionsmodells, daß für den lokal D-optimalen Versuchsplan mit p wiederholten Messungen die Likelihood-Konfidenzregionen (7) in etwa den linearisierten Regionen (6) entsprechen. Für den Original-Datensatz mit $n = 10$ verschiedenen, nicht geplanten Messungen hingegen weichen die Likelihood-Konfidenzregionen (7) stark von den linearisierten Regionen (6) ab; die lineare Approximation von $\eta(\theta)$ ist nicht geeignet. Der Nachteil vom D-Optimalitätskriterium und anderen an die lineare Versuchsplanung angelehnten

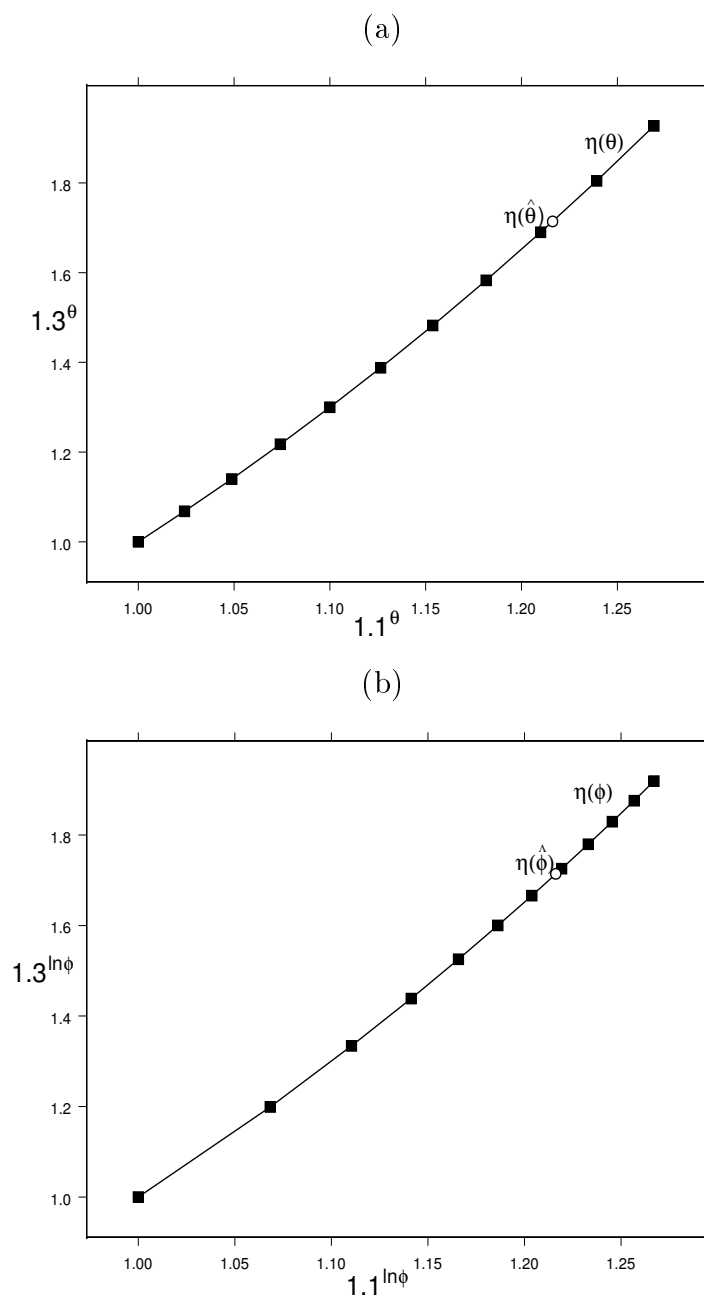


Abbildung 2: Eindimensionale Wirkungsflächen für $x_1 = 1.1, x_2 = 1.3$ von (a) Modell (2) und (b) Modell (3). Die Punkte in (a) entsprechen den θ -Werten $\theta = 0, 0.25, 0.5, \dots, 2.5$ ($\Delta\theta = 0.25$) und in (b) den ϕ -Werten $\phi = 1, 2, \dots, 12$ ($\Delta\phi = 1$).

Optimalitätskriterien ist neben der Abhängigkeit von θ die Abhängigkeit von der asymptotischen Normalverteilung von $\hat{\theta}$ und daher von der Adäquatheit von (8), der linearen Approximation der Wirkungshyperfläche. Wünschenswert ist es, daß Versuchspläne, die mit Hilfe eines bestimmten Optimalitätskriteriums und meist mit dem Ziel einer möglichst präzisen Schätzung der (oder von einigen der) unbekannt Parameter ermittelt werden, auch die Nichtlinearitätsmaße eines Modells verringern. Lokal D-optimale Versuchspläne führen im endlichen Stichprobenfall bei moderater Nichtlinearität (allgemeingültige Richtwerte gibt es keine) oft zu einer Verringerung der Nichtlinearität; sonst können D-optimale Versuchspläne hingegen auch zu Konfidenzregionen von unendlicher Größe (vgl. Hamilton and Watts, 1985) und zu einer Erhöhung der Nichtlinearität führen.

Für den Fall von stark ausgeprägter Nichtlinearität entwickelten einige Autoren Optimalitätskriterien, die weniger stark von der asymptotischen Theorie abhängen. Bates and Watts (1981) schlagen als Optimalitätskriterium z.B. vor, jenen Versuchsplan zu wählen, der die Parameter-Effekt-Nichtlinearität minimiert. Die Verringerung der Nichtlinearitätsmaße ist in diesem Kriterium explizites Ziel der Versuchsplanung. Hamilton and Watts (1985) entwickelten in Anlehnung an das D-Optimalitätskriterium ein quadratisches Versuchsplanungskriterium, das auf einer quadratischen Approximation des Volumens der exakten Konfidenzregion basiert. Sie empfehlen, bei stark ausgeprägter Nichtlinearität dieses Kriterium, ansonsten wegen der leichteren Berechenbarkeit das D-Optimalitätskriterium zu verwenden. Clarke (1980) entwickelte eine approximative Formel zweiter Ordnung für die Kovarianzmatrix von $\hat{\theta}$, die sowohl \ddot{F}^T als auch \ddot{F}^N enthält. Mit Hilfe dieser Formel erhielt Clarke eine Approximation zweiter Ordnung für den Kleinsten Quadratfehler ($KQF(\hat{\theta})$) von $\hat{\theta}$ (engl. *mean square error*, $MSE(\hat{\theta}) = KQF(\hat{\theta}) := E(\hat{\theta} - \theta)^2$). Als Optimalitätskriterium schlägt Clarke vor, die Approximation zweiter Ordnung des Kleinsten Quadratfehlers bezüglich des Versuchsplans zu minimieren. Das vorrangige Ziel der Versuchsplanung ist die präzise Schätzung der unbekannt Parameter; Clarke and Haines (1995) zeigen anhand eines ausgeprägt nichtlinearen Regressionsmodells, daß der bezüglich des angeführten Kriteriums (bzw. Modifikationen davon) ermittelte optimale Versuchsplan auch die Nichtlinearitätsmaße deutlich verringert. Pázman and Pronzato (1992) entwickelten ein Optimalitätskriterium, das an die Minimierung des Kleinsten Quadratfehlers angelehnt ist, jedoch auf einer nichtasymptotischen Wahrscheinlichkeitsdichte für die Schätzfunktion $\hat{\theta}$ basiert, die präziser als die Normalverteilung und für Intrinsische Nichtlinearität gleich Null exakt ist.

Clyde (1995) schlägt vor, jenen Versuchsplan zu wählen, der unter den Nebenbedingungen

$$\Gamma^T \sqrt{F} \leq k_T \quad \text{und} \quad \Gamma^N \sqrt{F} \leq k_N$$

lokal D-optimal ist. Sie berücksichtigt in ihrem Kriterium die Nichtlinearitätsmaße von Bates and Watts und verlangt von dem ermittelten optimalen Versuchsplan, daß er eine gewisse Grenze für die Nichtlinearitätsmaße von Bates and Watts nicht überschreitet, um so zu sichern, daß die auf der asymptotischen Theorie basierenden Schlüsse auch im endlichen Stichprobenfall valid sind. Als Optimalitätskriterien können auch andere Kriterien herangezogen werden.

Einen Überblick über die Versuchsplanung in der nichtlinearen Regression gibt

Tabelle 1: Blattfläche y_i (in m^2) von Ölpalmen in Abhängigkeit vom Alter x_i (in Jahren)

x_i	y_i
1	2.02
2	3.62
3	5.71
4	7.13
5	8.33
6	8.29
7	9.81
8	11.30
9	12.18
10	12.67
11	10.62
12	12.01

aus Rasch (1995)

Sedlacek (1996).

4 Fallstudie

Ziel der Fallstudie ist neben der Anwendung der oben angeführten Nichtlinearitätsmaße auf ein nichtlineares Regressionsmodell insbesondere zu zeigen, welche Effekte Parametertransformationen und Änderungen des Versuchsplans auf die Nichtlinearitätswerte haben. Es wird anhand des Beispiels gezeigt, daß die zur möglichst präzisen Schätzung der Parameter häufig verwendeten lokal D-optimalen Versuchspläne zu einer deutlichen Verringerung der Nichtlinearitätsmaße führen können.

Als Beispiel betrachten wir den Datensatz von Rasch (1995), in dem für 12 Ölpalmen die Blattfläche und das Alter angegeben sind (siehe Tabelle 1). Der nichtlineare Zusammenhang zwischen Alter (x) und Blattfläche (y) kann durch das Gompertzmodell

$$f(x_i; \boldsymbol{\theta}) = \theta_1 \exp[-\exp(\theta_2 - \theta_3 x_i)] \quad (10)$$

beschrieben werden (siehe Abbildung 3). Der KQS-Schätzer ergibt sich zu $\hat{\boldsymbol{\theta}} = (12.494, 0.915, 0.359)'$; $\hat{\sigma}^2 = 0.571$. Die maximale (relative) Parameter-Effekt-Krümmung berechnet sich zu $\Gamma^T = 1.510$ und die maximale (relative) Intrinsische Nichtlinearität zu $\Gamma^N = 0.188$ sowie die mittleren Quadrat-Nichtlinearitäten zu $\gamma_{RMS}^T = 0.685$ und $\gamma_{RMS}^N = 0.086$ (zur Implementierung siehe z.B. Venables and Ripley, 1997). Die relative Verzerrung der Parameter nach Box ist $\%B(\hat{\boldsymbol{\theta}}) = (0.823, 2.017, 1.380)'$. Mit $F(3, 9; 0.05) = 3.863$ ist $\Gamma_0 = 0.254$ und sowohl die durchschnittliche als auch die maximale Parameter-Effekt-Nichtlinearität daher als ausgeprägt anzusehen, während die Intrinsische Nichtlinearität vernachlässigbar ist. Ratkowsky (1983)

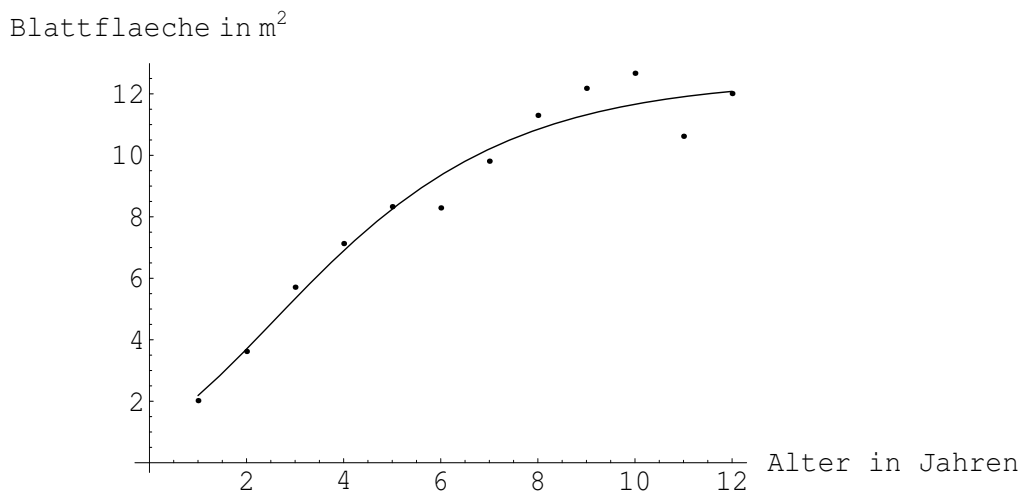


Abbildung 3: Blattfläche und Alter von Ölpalmen nach Rasch (1995) und das angepaßte Gompertz-Modell (9)

und Bates and Watts (1988) zeigen, daß die Intrinsische Nichtlinearität häufig vernachlässigbar ist, und die Parameter-Effekt-Nichtlinearität noch viel ausgeprägter als in diesem Beispiel sein kann, manchmal mit Werten über 50. Die Verzerrungen nach Box deuten darauf hin, daß vor allem θ_2 und θ_3 nichtlinear in ihrem Verhalten sind.

Um den Effekt einer geeigneten Reparametrisierung zu illustrieren, betrachten wir als Alternative das Gompertzmodell

$$f(x_i; \boldsymbol{\phi}) = \phi_1 \exp[\phi_2 \exp(\phi_3 x_i)]. \quad (11)$$

Diese Modellfunktion ergibt sich mit $\phi_1 = \theta_1$, $\phi_2 = -\exp^{\theta_2}$ und $\phi_3 = -\theta_3$ aus (10). Für die KQS-Schätzer erhält man $\hat{\boldsymbol{\phi}} = (12.494, -2.498, -0.359)'$. Die Intrinsische Nichtlinearität bleibt wie oben erwähnt gleich, und die Parameter-Effekt-Nichtlinearität läßt sich zu $\Gamma^T = 1.390$ (bzw. $\gamma_{RMS}^T = 0.685$) berechnen. Die Reparametrisierung hat somit zu keiner wirklichen Verbesserung geführt. Die Verzerrungen nach Box bestätigen das: $\%B(\hat{\boldsymbol{\phi}}) = (0.823, 3.404, 1.380)'$; ϕ_2 zeigt nach dieser Berechnung trotz etwas kleinerem Γ^T ein noch stärker nichtlineares Verhalten als θ_2 . Dies läßt sich darauf zurückzuführen, daß die Verzerrung der Parameter nicht von der maximalen relativen Parameter-Effekt-Nichtlinearität abhängt, sondern von $\ddot{\mathbf{F}}^T$.

Für eine weitere Modellfunktion des Gompertz-Modells,

$$f(x_i; \boldsymbol{\psi}) = \exp[\psi_1 - \psi_2 \psi_3^{x_i}], \quad (12)$$

ergeben sich $\Gamma^T = 1.027$ (bzw. $\gamma_{RMS}^T = 0.577$) und $\%B(\hat{\boldsymbol{\psi}}) = (0.267, 3.404, -0.272)'$.

Die Parameter-Effekt-Nichtlinearität konnte durch Reparametrisierung von ϕ_3 reduziert werden; die weiterhin ausgeprägte Nichtlinearität ist nun größtenteils auf den zweiten, unverändert gebliebenen Parameter zurückzuführen.

Rasch (1995) ermittelte für das Gompertz Modell (11) den lokal D-optimalen Versuchsplan X_D^* , wobei für die wahren Parameter die Parameterschätzer $\hat{\phi} = (12.494, -2.498, -0.359)'$ stehen:

$$X_D^* = \begin{pmatrix} 1 & 4.953 & 12 \\ 4 & 4 & 4 \end{pmatrix}. \quad (13)$$

Er zeigt, daß der lokal D-optimale Versuchsplan robust gegenüber Veränderungen der Parameter ϕ ist. Der lokal D-optimale Versuchsplan besteht aus $n = p$ verschiedenen Beobachtungen, die wiederholt gemessen werden können. Viele Autoren haben sich eingehend mit den Vorteilen von Versuchsplänen mit $n = p$ verschiedenen Punkten beschäftigt (siehe z.B. Box and Lucas, 1959, und Box, 1970). So ist z.B. die Intrinsische Nichtlinearität für diese Versuchspläne gleich Null. Allerdings muß darauf hingewiesen werden, daß es für lediglich p verschiedene Versuchspunkte kaum möglich ist, zu überprüfen, wie gut sich ein gewähltes Modell den Daten anpaßt.

Die folgenden Nichtlinearitätsberechnungen der veränderten Versuchspläne basieren auf der Annahme, daß die Parameterschätzer $\hat{\phi} = (12.494, -2.498, -0.359)'$ und $\hat{\sigma}^2 = 0.571$ die wahren Parameter sind.

Für den lokal D-optimalen Versuchsplan (13) wurden die maximalen Nichtlinearitäten nach Bates and Watts und die relative Verzerrung des KQS-Schätzers berechnet: $\Gamma^T = 0.599$, $\Gamma^N = 0$ und $\%B(\hat{\phi}) = (0.347, 1.594, 0.743)'$. Der lokal D-optimale Versuchsplan führt somit zu einer deutlichen Verringerung aller Nichtlinearitätswerte. Da es oft Situationen gibt, in denen keine wiederholten Messungen möglich sind, betrachten wir zur Illustration den an den lokal D-optimalen Versuchsplan angelehnten wiederholungsfreien Versuchsplan $X_1 = (0.5, 1, 1.5, 2, 4, 4.5, 5, 5.5, 11, 11.5, 12, 12.5)$. Wir erhalten $\Gamma^T = 0.662$, $\Gamma^N = 0.078$ und $\%B = (0.433, 1.883, 0.784)$. Die Werte entsprechen in etwa den Ergebnissen für den lokal D-optimalen Versuchsplan und sind deutlich kleiner als die Nichtlinearitäten für das gleiche Modell (11) und $X = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$.

Zur weiteren Illustration der Effekte des Versuchsplans nehmen wir an, daß die Blattfläche von Ölpalmen nur bis zum Alter von 8 Jahren gemessen werden konnte, und betrachten den Versuchsplan $X_2 = (1, 2, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 8)$, der somit den vermuteten Wertebereich für das Alter von Ölpalmen (1-12 Jahre) nicht vollständig abdeckt. Die Nichtlinearitätsmessungen ergeben $\Gamma^T = 4.747$, $\Gamma^N = 0.198$, und die relativen Verzerrungen der einzelnen Parameter liegen alle über 1%. Vor allem die Parameter-Effekt-Nichtlinearität ist für diesen Versuchsplan weithaus höher als für den ursprünglichen Versuchsplan aus Tabelle 1 und für den lokal D-optimalen Versuchsplan (13).

Tabelle 2 zeigt zusammenfassend die Ergebnisse der Fallstudie. Für alle Werte zeigt das Modell mit dem lokal D-optimalen Versuchsplan und dem daraus entstandenen Versuchsplan X_1 das geringste nichtlineare Verhalten.

Tabelle 2: Nichtlinearitätswerte

	Γ^N	Γ^T	$\%B$
Gompertz-Modell (10)	0.188	1.510	(0.823, 2.017, 1.38)'
Modellfunktion (11)	0.188	1.390	(0.823, 3.404, 1.38)'
Modellfunktion (12)	0.188	1.027	(0.267, 3.404, -0.272)'
lokal D-optimaler Versuchsplan	0	0.599	(0.347, 1.594, 0.743)'
Versuchsplan X_1	0.078	0.662	(0.433, 1.883, 0.784)
Versuchsplan X_2	0.198	4.747	(4.159, 5.652, 1.9)'

5 Zusammenfassung und Schlußfolgerungen

Es ist wünschenswert, daß statistische Aussagen, die auf der asymptotischen Theorie aufbauen und unter der Voraussetzung, daß die Schätzfunktion $\hat{\theta}$ approximativ nach $N_p(\theta^*, \sigma^2(\dot{\mathbf{F}}'\dot{\mathbf{F}})^{-1})$ verteilt ist, getroffen werden, im endlichen Stichprobenfall zu validen Ergebnissen führen. Die Nichtlinearitätsmaße nach Bates and Watts sind geeignete Indikatoren, um das nichtlineare Verhalten eines Regressionsmodells zu quantifizieren. Die Verzerrung nach Box gibt zusätzlich Anhaltspunkte, welche Parameterkomponenten am meisten zur Nichtlinearität beitragen. Die Nichtlinearitätsmaße lassen Rückschlüsse zu, in welchem Maße statistische Aussagen, die auf der linearen Approximation der gekrümmten Hyperfläche der Erwartungswerte $\eta(\theta)$ basieren, valid sind.

Die Parameter-Effekt-Nichtlinearität hängt von der Parametrisierung des Modells ab; passende Reparametrisierungen können häufig zu einer deutlichen Reduzierung der Parameter-Effekt-Nichtlinearität führen (Ratkowsky, 1983). Die Intrinsische Nichtlinearität hängt neben der Gestalt des Modells vom verwendeten Versuchsplan ab. Ziel von Versuchsplanungen ist es häufig, jenen Versuchsplan zu ermitteln, der eine möglichst präzise Schätzung der Parameter zuläßt. Wünschenswert ist es, daß die bezüglich eines bestimmten Optimalitätskriteriums ermittelten Versuchspläne wie etwa lokal D-optimale Versuchspläne auch die Nichtlinearität eines Modells verringern. Sind die Nichtlinearitätswerte eines Modells moderat, liefern die an die lineare Versuchsplanung angelehnten Versuchsplanungen zur möglichst präzisen Schätzung der unbekannt Parameter valide Ergebnisse und führen auch häufig zu einer Verringerung der Nichtlinearitätsmaße. Im Falle von ausgeprägter Nichtlinearität sind aufwendigere Verfahren zur Versuchsplanung notwendig, die die Nichtlinearität eines Modells berücksichtigen und trotz hoher Nichtlinearität zu Versuchsplänen führen, die eine möglichst präzise Schätzung der Parameter ermöglichen, und häufig zusätzlich eine Verringerung der Nichtlinearität bewirken können.

Die Fallstudie demonstriert den Effekt von Veränderungen des Versuchsplans auf die Nichtlinearität. Sie zeigt, daß lokal D-optimale Versuchspläne die Nichtlinearität deutlich verringern können. Die Verkleinerung des Wertebereichs des Versuchsplans hingegen führt zu einer deutlichen Erhöhung der Nichtlinearitätswerte.

6 Danksagung

Ich danke dem Referee für seine sehr konkreten und hilfreichen Hinweise und den Professoren Univ.Prof. Dr. Peter Hackl und a.o.Univ.Prof. Dr. Werner Müller, Abteilung für Wirtschaftsstatistik an der Wirtschaftsuniversität Wien, für ihre wertvolle Unterstützung.

Literatur

- D.M. Bates and D.G. Watts. Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 42:1–25, 1980.
- D.M. Bates and D.G. Watts. Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Annals of Statistics*, 9:1152–1167, 1981.
- D.M. Bates and D.G. Watts. *Nonlinear Regression Analysis and its Applications*. Wiley & Sons, New York, 1988.
- E.M.L. Beale. Confidence regions in non-linear estimation (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 22:41–88, 1960.
- G.E.P. Box and H.L. Lucas. Design of experiments in nonlinear situations. *Biometrika*, 47:77–99, 1959.
- M.J. Box. Bias in nonlinear estimation (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 33:171–201, 1971.
- G.P.Y. Clarke. Moments of the least squares estimators in a non-linear regression model. *Journal of the Royal Statistical Society, Ser. B*, 42:227–237, 1980.
- G.P.Y. Clarke and L.M. Haines. Optimal design for models incorporating the richards function. In G.U.H. Seeber, B.J. Francis, R. Hatzinger, and G. Steckelberger, editors, *Statistical Modelling - Proceedings of the 10th International Workshop on Statistical Modelling*, pages 61–66, 1995.
- M.A. Clyde. Bayesian design for approximate normality. In C.P. Kitsos and W.G. Müller, editors, *MODA 4 - Advances in Model-Oriented Data Analysis*, pages 25–37, 1995.
- D.C. Hamilton and D.G. Watts. A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics*, 27:241–250, 1985.
- A. Pázman. *Nonlinear Statistical Models*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1993.
- A. Pázman and L. Pronzato. Nonlinear experimental design based on the distribution of estimators. *Journal of Statistical Planning and Inference*, 33:385–403, 1992.

- D. Rasch. The robustness against parameter variation of exact locally optimum designs in nonlinear regression - a case study. *Computational Statistics and Data Analysis*, 20:441–453, 1995.
- D.A. Ratkowsky. *Nonlinear Regression Modeling*. Marcel Dekker, New York, 1983.
- G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. Wiley & Sons, New York, 1989.
- G. Sedlacek. *Nichtlineare Regression und Versuchsplanung*. Diplomarbeit, Technische Universität Wien, 1996.
- W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer, New York, 1997.

Adresse des Autors:

Dipl.-Ing. Günther Sedlacek
Abteilung für Wirtschaftsstatistik
Wirtschaftsuniversität Wien
Augasse 2-6, A-1090 Wien
Tel.: 43 1 31336/5229 Fax: /711
E-mail: guenther.sedlacek@isis.wu-wien.ac.at