Schätzung partiell identifizierter Parameter in generalisierten linearen Modellen mit Intervalldaten

Vortrag zur Masterarbeit

Michael J. Seitz

Betreuer: Prof. Dr. Thomas Augustin

Arbeitsgruppe Method(olog)ische Grundlagen der Statistik und ihre Anwendungen
Institut für Statistik
Ludwig-Maximilians-Universität München

29. Oktober 2012

Übersicht des Vortrags

- Partielle Identifizierbarkeit und Intervalldaten
- Generalisierte lineare Regression
- Schätzung der partiell identifizierten Parameter
- Analytische Lösung für einen Spezialfall
- Iterative Optimierung der Score-Funktion
- Numerische und heuristische Ansätze
- Verwendete Optimierungsverfahren
- Simulationsbeispiele
- Zusammenfassung
- Ausblick

Partielle Identifizierbarkeit

- Bei partieller Identifizierbarkeit des Parameters wird die Schätzung eines exakten Parameters häufig durch rigide Annahmen ermöglicht.
- Neuere Theorien versuchen diese Annahmen aufzulockern oder ganz ohne zusätzliche Annahmen auszukommen
  [Manski, 2003, Beresteanu und Molinari, 2012, Moon und Schorfheide, 2012].
- Die Unsicherheit in der Beobachtung wird somit auf den Schätzer übertragen.
- Das Ergebnis ist kein Punktschätzer, sondern ein Identifizierungsbereich für den Parameter.

Intervalldaten

- Bei Intervalldaten geht man davon aus, dass die Beobachtungen in Intervallen liegen, von denen die Ober- und Untergrenzen bekannt sind:

$$\mathfrak{x} = (\mathfrak{x}_1, \ldots, \mathfrak{x}_n)$$
$$\mathfrak{y} = (\mathfrak{y}_1, \ldots, \mathfrak{y}_n)$$

$$\text{mit } \mathfrak{x}_i = [\underline{x}_i, \overline{x}_i]$$
$$\text{und } \mathfrak{y}_i = [\underline{y}_i, \overline{y}_i]$$

- Wie die Werte in diesen Intervallen verteilt sind ist nicht bekannt.

Intervalldaten

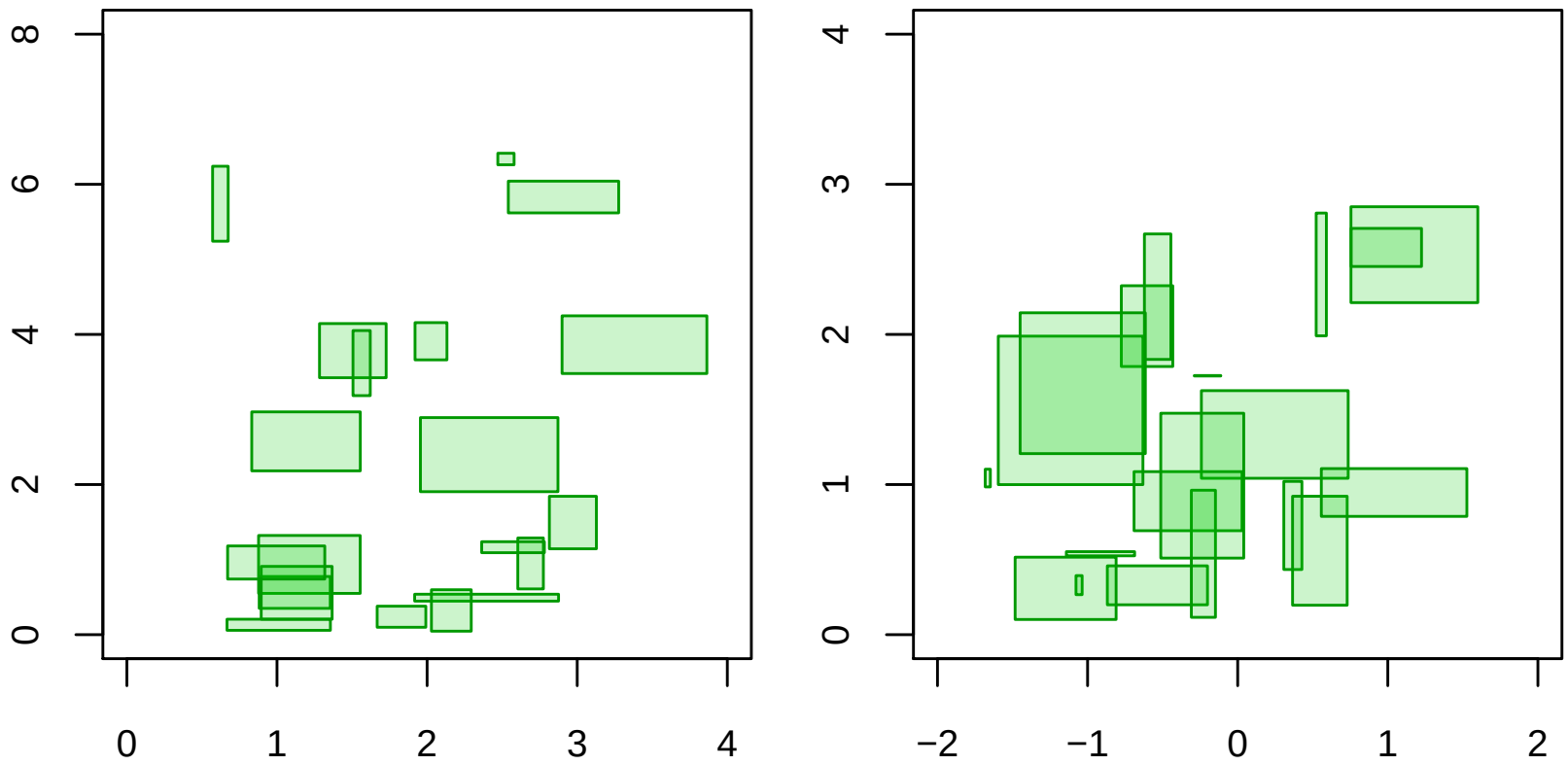


Abbildung: Simulierte Intervalldaten.

Generalisierte lineare Regression

- Modellannahmen:
  - Die abhängige Variable y stammt aus einer Exponentialfamilie.
  - Der Erwartungswert wird mit einer Link-Funktion und dem linearen Prädiktor modelliert:
$$E(y|\boldsymbol{x}) = h(\boldsymbol{x}^T\beta).$$
- Schätzung mit der Maximum- oder Quasi-Likelihood-Methode.

➥ Mit skalaren Daten muss ein Optimierungsproblem ohne Nebenbedingungen gelöst werden.

Schätzung der partiell identifizierten Parameter

- Ziel ist es, das Intervall $I(\beta_k) = (\underline{\beta}_k, \overline{\beta}_k)$ zu bestimmen, dass alle *zulässigen* Parameterschätzer enthält.
- Ein Parameterschätzer $\hat{\beta}_k$ ist zulässig, wenn es Punkte aus den Intervalldaten gibt, die zu diesem Schätzer führen.
- Als Kriterium für die Zulässigkeit wird eine Schätzfunktion $\Psi(\mathbf{x}, \mathbf{y}, \beta) = 0$ verwendet [Augustin, 2012].
- Im Weiteren ist Ψ die Score-Funktion und die Gleicheitsbedingung das Maximum-Likelihood-Kriterium.

➥ Man erhält ein Optimierungsproblem mit Nebenbedingungen.

Schätzung der partiell identifizierten Parameter

- Lineare Zielfunktion mit nicht-linearen Gleichheitsnebenbedingungen und Box-Constraints [Augustin, 2012]:

$$\beta_k \to \min \ / \ \max, \qquad k = 1, \ldots, q \tag{1}$$

mit den Nebenbedingungen

$$\begin{aligned} \Psi(\mathbf{x}, \mathbf{y}, \beta) &= 0 \\ x_i &\in \mathfrak{X}_i, \qquad i = 1, \ldots, n \\ y_i &\in \mathfrak{Y}_i, \qquad i = 1, \ldots, n \end{aligned}$$

Schätzung der partiell identifizierten Parameter

- Nicht-lineare Zielfunktion mit Box-Constraints [Augustin, 2012]:

$$\beta_k \pm \rho \left(\Psi(\mathbf{x}, \mathbf{y}, \beta)\right)^2 \rightarrow \min \ / \ \max \tag{2}$$

mit den Nebenbedingungen

$$\begin{aligned} x_i &\in \mathfrak{X}_i, \qquad i = 1, \ldots, n \\ y_i &\in \mathfrak{Y}_i, \qquad i = 1, \ldots, n \end{aligned}$$

➥ Grundidee ist also, die Punkte aus den Intervallen der Daten auszuwählen, die zu den extremsten Parameterschätzern führen.

Analytische Lösung für einen Spezialfall

- Gegeben seien die unabhängige Variable

$$x = \underline{x} = \overline{x}.$$

und die abhängige Variable

$$y = \underline{y} + \delta$$

wobei $0 \leq \delta \leq \overline{y} - \underline{y}$.

➥ Im linearen Modell

$$E(y) = \beta_0 + x\beta_1$$

kann $I(\beta_1)$ analytisch bestimmt werden, wenn die x-Werte als Skalare und nur die y-Werte als Intervalle gegeben sind [Rohwer und Pötter, 2001].

Analytische Lösung für einen Spezialfall

- Die zulässigen Parameterschätzer sind gegeben durch

$$\hat{\beta}_1(\delta_1, \ldots, \delta_n) = \frac{n \sum_{i=1}^n x_i(\underline{y}_i + \delta_i) - \sum_{i=1}^n x_i \sum_{i=1}^n (\underline{y}_i + \delta_i)}{n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i} \tag{3}$$

$$\propto \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \delta_i \tag{4}$$

➥ Für das Minimum von β_1 muss also gelten:

$$\delta_i = \begin{cases} \overline{y} - \underline{y} & \text{wenn } x_i < \frac{1}{n} \sum_{j=1}^n x_j \\ 0 & \text{sonst} \end{cases}$$

[Rohwer und Pötter, 2001]

Analytische Lösung für einen Spezialfall

- Das Maximum wird analog bestimmt, wobei die inverse Regel angewandt wird.
- Die Intervallgrenzen werden anschließend durch klassische Schätzung mit den jeweiligen Daten berechnet.
- Dabei wird nur der Identifizierungsbereich für β_1 bestimmt.
- Ähnlich dem Ansatz in [Rohwer und Pötter, 2001] kann dieser aber auch für β_0 bestimmt werden.
- Das Modell ohne Intercept mit

$$E(y) = x\beta_1$$

kann ebenfalls ähnlich gelöst werden.

Analytische Lösung für einen Spezialfall

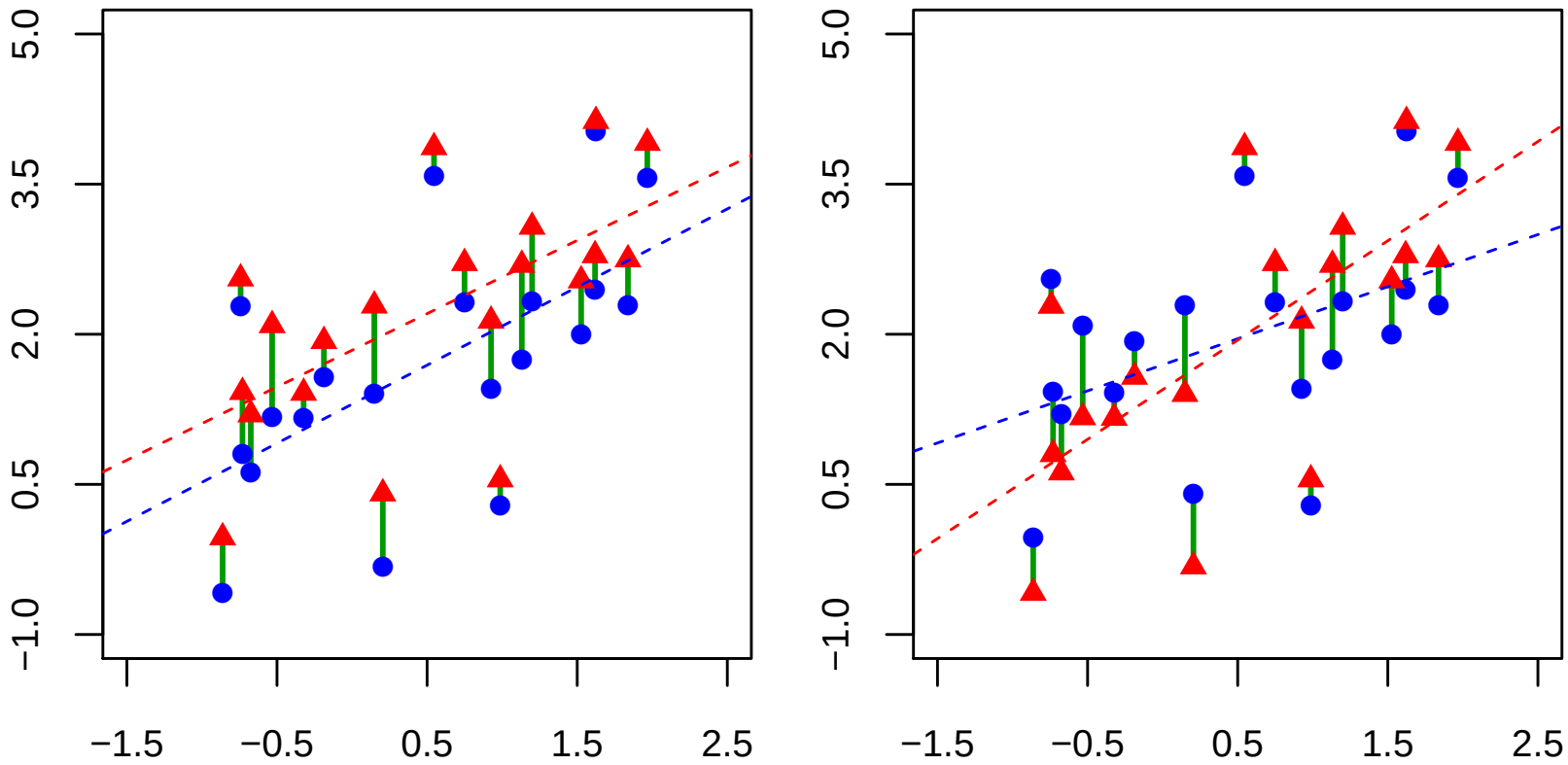


Abbildung: Grenzen der Regressionsgeraden bei gegebenen Intervalldaten für die Parameterschätzer $\hat{\beta}_0$ links und $\hat{\beta}_1$ rechts.

Iterative Optimierung der Score-Funktion

- Ist der Parameter β eindimensional, kann die Score-Funktion als Summe dargestellt werden [Augustin, 2012]:

$$\Psi(\beta; \boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \Psi_i(\beta; x_i, y_i) \tag{5}$$

➥ Die einzelnen Summanden hängen jeweils nur von dem Parameter und einer Beobachtung ab.

Iterative Optimierung der Score-Funktion

- Ist Ψ in β strikt monoton fallend, so gilt

$$\Psi(\hat{\beta}^{(t)}; \boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) > \Psi(\hat{\beta}^{(t)}; \boldsymbol{x}^{(t+1)}, \boldsymbol{y}^{(t+1)}) \tag{6}$$

$$\Rightarrow \hat{\beta}^{(t)} > \hat{\beta}^{(t+1)} \tag{7}$$

mit

$$\begin{aligned} \Psi(\hat{\beta}^{(t)}; \boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}) &= 0, \\ \Psi(\hat{\beta}^{(t+1)}; \boldsymbol{x}^{(t+1)}, \boldsymbol{y}^{(t+1)}) &= 0, \\ \boldsymbol{x}^{(m)}, \boldsymbol{x}^{(m+1)} &\in \mathfrak{X}, \\ \boldsymbol{y}^{(m)}, \boldsymbol{y}^{(m+1)} &\in \mathfrak{Y}. \end{aligned}$$

[Augustin, 2012]

Iterative Optimierung der Score-Funktion

- Durch Minimierung von Ψ über $\boldsymbol{x}$ und $\boldsymbol{y}$ kann also ein kleinerer zulässiger Schätzer $\hat{\beta}$ gefunden werden.
- Durch Aufteilung der Score-Funktion in einzelne Summanden kann das Minimum effizient analytisch bestimmt werden.

Iterative Optimierung der Score-Funktion

- Es gilt, wenn

$$\Psi(\hat{\beta}^{(m)}; \boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)}) = 0$$

und

$$\nexists\ (\boldsymbol{x}^{(m+1)}, \boldsymbol{y}^{(m+1)})$$

mit

$$\Psi(\hat{\beta}^{(m)}; \boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)}) > \Psi(\hat{\beta}^{(m)}; \boldsymbol{x}^{(m+1)}, \boldsymbol{y}^{(m+1)}),$$

wobei

$$\boldsymbol{x}^{(m)}, \boldsymbol{x}^{(m+1)} \in \mathfrak{X},$$

$$\boldsymbol{y}^{(m)}, \boldsymbol{y}^{(m+1)} \in \mathfrak{Y},$$

dann

$$\Rightarrow \hat{\beta}^{(m)} = \underline{\beta}.$$

Iterative Optimierung der Score-Funktion

- Es kann also iterativ so lange ein kleineres $\hat{\beta}$ bestimmt werden, bis keine Veränderung mehr auftritt und somit das Minimum $\underline{\beta}$ erreicht wurde.
- Das Maximum lässt sich analog bestimmen.

Iterative Optimierung der Score-Funktion

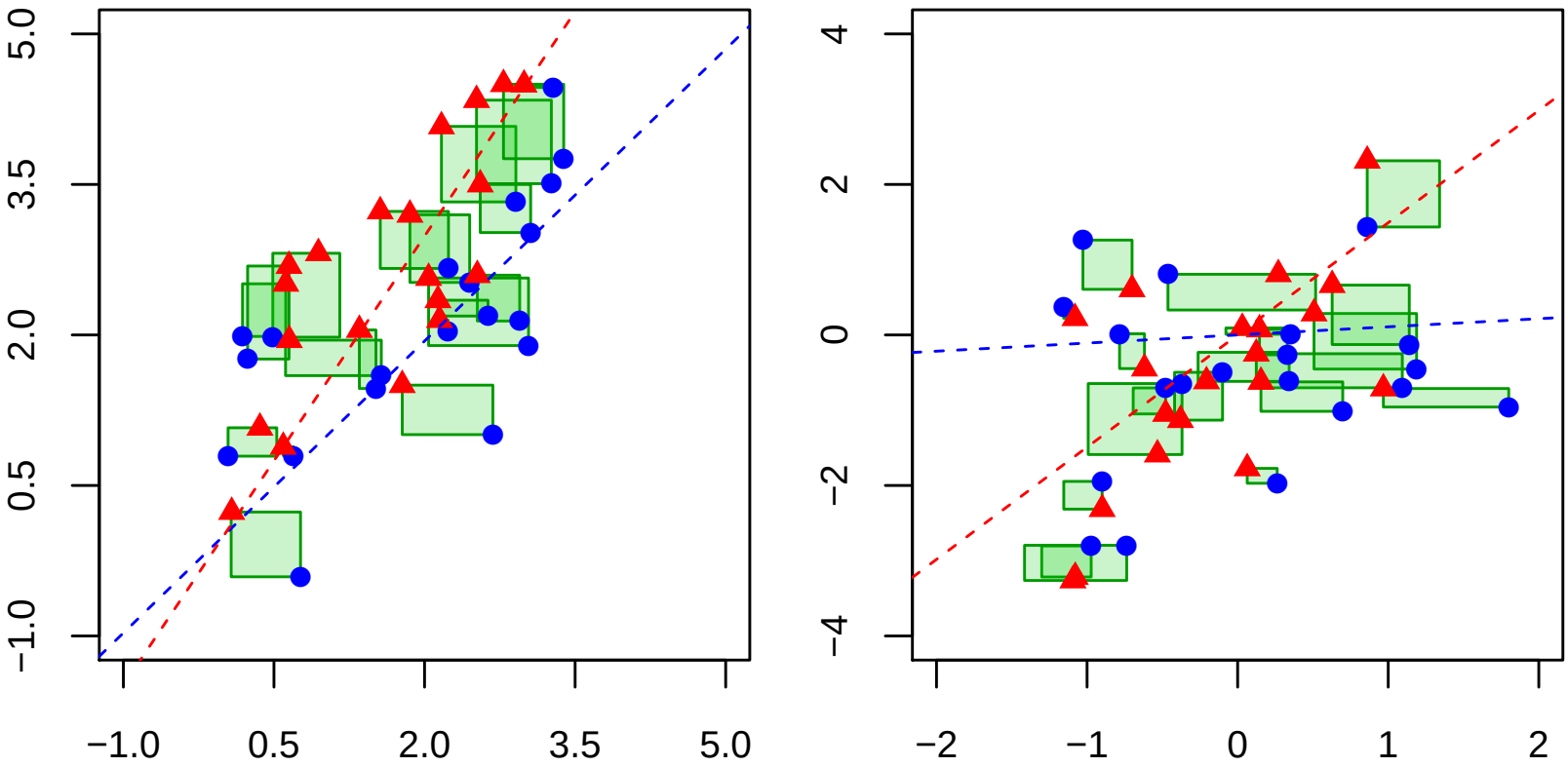


Abbildung: Grenzen der Regressionsgeraden bei gegebenen Intervalldaten für den Parameterschätzer $\hat{\beta}$ für zwei Simulationsbeispiele. Es wird ein lineares Modell ohne Intercept angenommen.

Iterative Optimierung der Score-Funktion

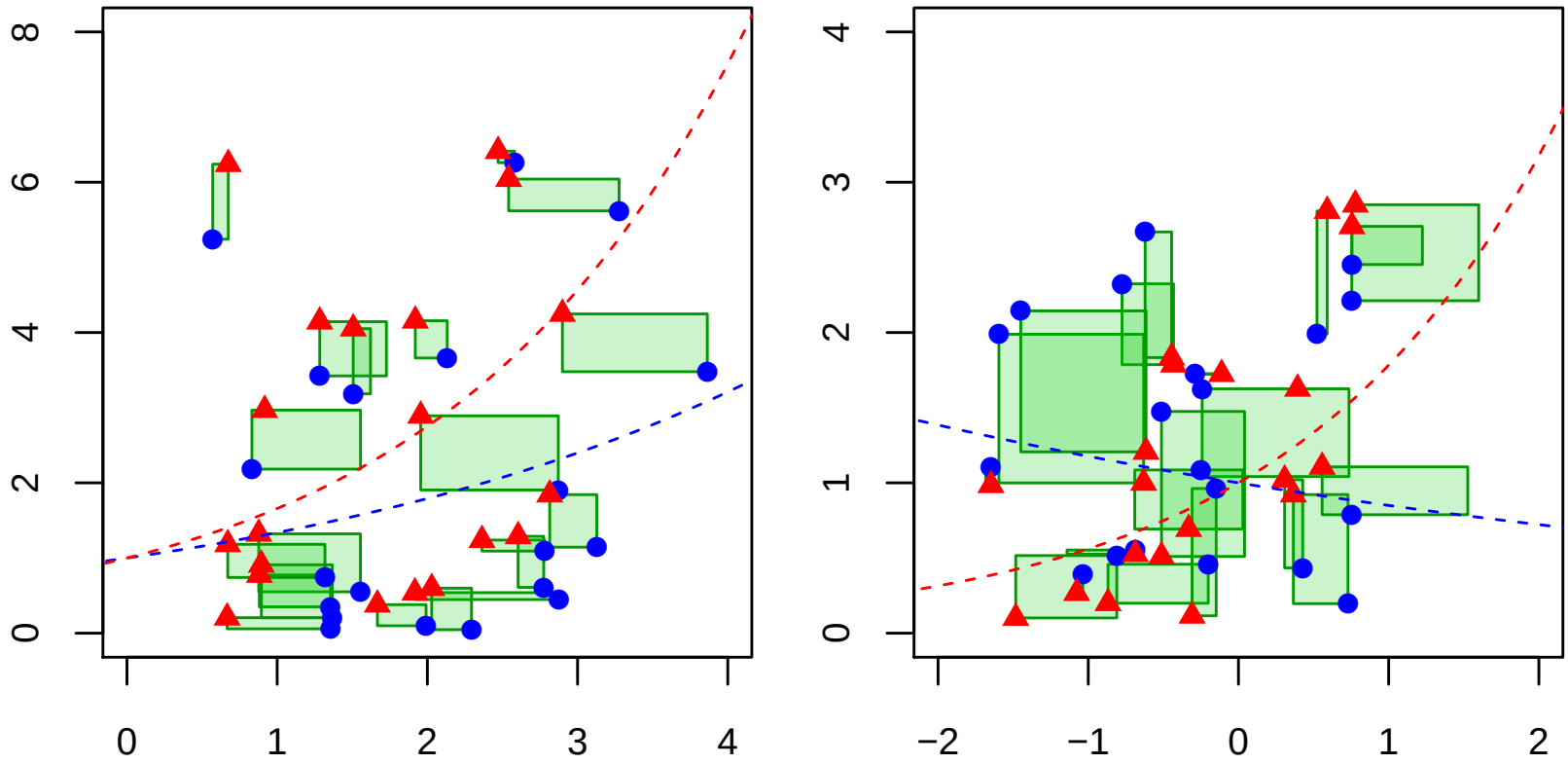


Abbildung: Grenzen der Regressionskurven bei gegebenen Intervalldaten für den Parameterschätzer $\hat{\beta}$ für zwei Simulationsbeispiele. Es wird ein generalisiertes lineares Modell mit Exponentialverteilung und log-Link angenommen.

Numerische und heuristische Ansätze (1)

Optimierung des Parameters mit Strafterm für die Nebenbedingung der Score-Funktion und Box-Constraints:

$$\min_{\beta,\boldsymbol{x},\boldsymbol{y}} / \max_{\beta,\boldsymbol{x},\boldsymbol{y}} \; \beta_k \pm \rho \left(\Psi(\boldsymbol{x}, \boldsymbol{y}, \beta)\right)^2 .$$

- Einsatz eines numerischen Optimierungsverfahrens (der Gradient kann analytisch bestimmt werden).
- Dabei hat die Wahl des Gewichts ρ einen großen Einfluss auf das Ergebnis: Zu große Werte führen zu nicht optimalen, und zu kleine zu unzulässigen Schätzwerten.

➥ Erneute Schätzung der Parameter mit gegebenen Punkten um zulässigen Schätzer zu erhalten.

➥ Mehrmalige Optimierung mit sukzessiver Erhöhung von ρ.

Numerische und heuristische Ansätze (2)

Direkte Optimierung des Parameterschätzers mit Box-Constraints:

$$\min_{\boldsymbol{x},\boldsymbol{y}} / \max_{\boldsymbol{x},\boldsymbol{y}} \ \hat{\beta}_k(\boldsymbol{x}, \boldsymbol{y}).$$

- Der Parameterschätzer $\hat{\beta}_k$ liegt entweder als geschlossene Form vor (lineares Modell) oder muss numerisch bestimmt werden (generalisiertes lineares Modell).
- Im zweiten Fall ist die numerische Schätzung deutlich aufwendiger und der Gradient der Zielfunktion kann nicht bestimmt werden.

➥ Für die Bestimmung der Parameterschätzer können vorhandene, sehr effiziente numerische Verfahren eingesetzt oder bekannte geschlossene Formen verwendet werden.

Numerische und heuristische Ansätze (3)

Sequentielle, unabhängige Überprüfung der Ecken in den Intervallen der Daten.

- Bei konvexen und konkaven Zielfunktionen mit Box-Constraints liegen die Extrema häufig in den Ecken.
- Berechnung des Parameterschätzers für alle Ecken einer Beobachtung – die anderen Beobachtungen werden festgehalten.
- Effizient, da die Dimension der Zielfunktion mit der Anzahl der Beobachtungen nicht wächst.
- Wiederholtes Durchlaufen aller Ecken, bis keine Veränderung mehr auftritt.

Numerische und heuristische Ansätze (4)

Sequentielle, unabhängige Suche ausgehend von den Ecken in den Intervallen der Daten:

$$\min_{x_i, y_i} / \max_{x_i, y_i} \hat{\beta}_k(\boldsymbol{x}, \boldsymbol{y})$$

- Für jede Beobachtung i wird ausgehend von allen 2^{p+1} Ecken das Extremum gesucht, während die Punkte für die anderen Beobachtungen gleich bleiben.
- Effizient, da die Dimension der Zielfunktion mit der Anzahl der Beobachtungen nicht wächst.
- Wiederholtes Durchlaufen, bis keine Veränderung mehr auftritt.

Numerische und heuristische Ansätze

- Für die direkte Optimierung und die Optimierung mit Strafterm wächst die Dimension der Zielfunktion mit jeder weiteren Beobachtung.
- Alle Beobachtungen werden gleichzeitig betrachtet.
- Trotzdem kann es sein, dass nur ein lokales Extremum gefunden wird.
- Bei der Überprüfung der Ecken und der Suche ausgehend von den Ecken werden die einzelnen Beobachtungen nacheinander und nicht gleichzeitig betrachtet.
- Dimension der Zielfunktion wächst nicht mit weiteren Beobachtungen.
- Durch Neustart in allen Ecken kann unter Umständen aus einem lokalem Extremum entkommen werden.
- Gefahr nur ein lokales Extremum zu finden bleibt aber dennoch bestehen.

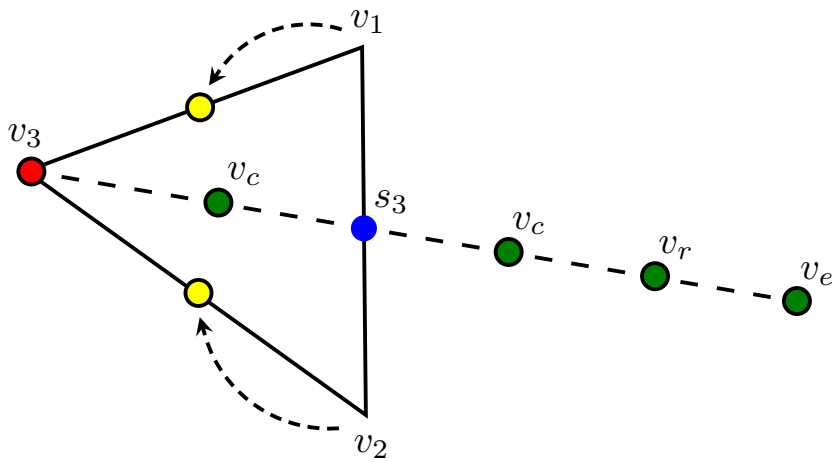
Verwendete Optimierungsverfahren

- Das R-Modul `optim` [R Development Core Team, 2012] bietet einige Optimierungsverfahren an.
- Auf Grund der hohen Dimensionen der Zielfunktionen werden effiziente Optimierungsverfahren benötigt.
- Ein Grid oder häufiger Neustart an verschiedenen Stellen ist nur sehr eingeschränkt möglich.
- Die Zielfunktion wird mit weiteren Beobachtungen und mehr unabhängigen Einflussgrößen komplexer.
- Hier Beschränkung auf zwei Verfahren.

Verwendete Optimierungsverfahren (1)

Verfahren nach Nelder und Mead [Nelder und Mead, 1965]

- Das Verfahren nach Nelder und Mead basiert auf der wiederholten Manipulation eines Simplex ohne dabei den Gradienten der Zielfunktion zu verwenden.



- Es folgt in gewisser Weise einer Abstiegsrichtung und liefert unter Umständen nur ein lokales Extremum.
- Grundsätzlich für die Optimierung ohne Restriktionen geeignet, kann aber mit einer Barrier-Methode kombiniert werden [Lange, 2010].

Verwendete Optimierungsverfahren (2)

L-BFGS-B [Byrd et al., 1995]

- Verfahren für die Optimierung mit Box-Constraints unter Verwendung des Gradienten.
- Keine Erweiterung des L-BFGS-Verfahrens [Nocedal, 1980], sondern verwendet die Approximation dieser Verfahren zur Schätzung der Hesse-Matrix.
- Dabei quadratische Approximation an der Stelle des aktuellen Punktes.
- Verwendet eine Abstiegsrichtung um den Punkt der nächsten Iteration zu bestimmen.
- Projektion in den zulässigen Bereich.

Simulationsbeispiele

- Beschränkung auf das lineare Modell und das generalisierte lineare Modell mit Exponentialverteilung und log-Link.
- Es werden nur lineare Prädiktoren mit einer unabhängigen Variable (mit und ohne Intercept) untersucht.
- Vergleich der Intervalle des Parameterschätzers und der ausgewählten Punkte aus den Datenintervallen für die verschiedenen Verfahren.

Simulationsbeispiele: Simulationsmodelle (ein Beispiel)

- Daten werden aus Gleich- und Normalverteilungen (bzw. Exponentialverteilungen) simuliert:

$$\underline{x} \sim U(0,3),\ \underline{y} \sim N(\underline{x},1).$$

- Der wahre Zusammenhang besteht zwischen den Untergrenzen $\underline{x}$ und $\underline{y}$:

$$\underline{y} = \underline{x}\beta + \epsilon$$

  mit $\beta = 1$.
- Für die Obergrenzen gilt

$$\overline{x} = \underline{x} + u_1,\ \overline{y} = \underline{y} + u_2$$

  mit $u_1, u_2 \sim U(0,1)$.
- Verschiedene Stichprobenumfänge von $n = 10, 20, 1000$.

Simulationsbeispiele: Ergebnisse Modelle ohne Intercept

- Die Lösung kann zuverlässig durch iterative analytische Optimierung der Score-Funktion bestimmt werden.
- Die direkte Optimierung und die Optimierung mit Strafterm liefern teilweise nur lokale Extrema, die aber nur in einigen wenigen Beobachtungen von dem globalen Extremum abweichen.
- Nur das Verfahren L-BFGS-B konvergiert dabei sinnvoll zu einem Extremum.
- Das Verfahren nach Nelder und Mead ist deutlich langsamer und lieferte in allen Fällen deutlich schlechtere Ergebnisse.
- Die wiederholte Suche ausgehend von den Ecken liefert in allen untersuchten Fällen die richtige Lösung.

Simulationsbeispiele: Unterschiede zur Referenzlösung

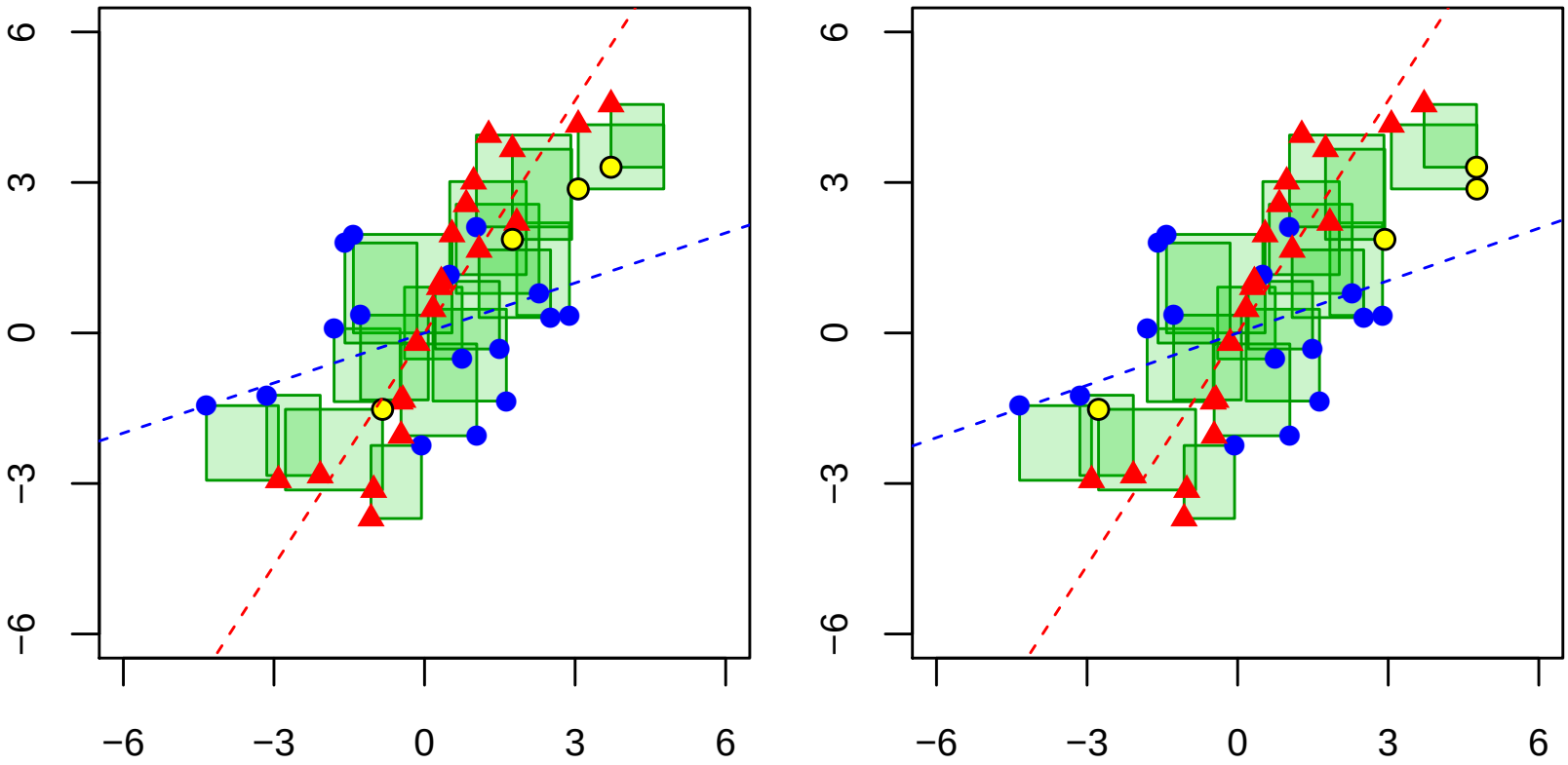


Abbildung: Unterschiede der Ergebnisse für die Iterative Optimierung der Score-Funktion als Refernzlösung (links) und der direkten Optimierung (rechts) in einem linearen Modell ohne Intercept.

Simulationsbeispiele: Ergebnisse Modelle mit Intercept

- Es gibt kein Refernzverfahren zur Bestimmung der Lösung.
- Die direkte Optimierung und die Optimierung mit Strafterm liefern teilweise nur lokale Extrema, die aber wieder nur in einigen wenigen Beobachtungen von dem globalen Extremum abweichen.
- Für die numerischen Optimierungsverfahren gilt das gleiche wie in den Modellen ohne Intercept.
- Die wiederholte Suche ausgehend von den Ecken kann unter Umständen das globale Extremum nicht finden (dies konnte aber nicht direkt beobachtet werden).

Simulationsbeispiele: Unterschiede

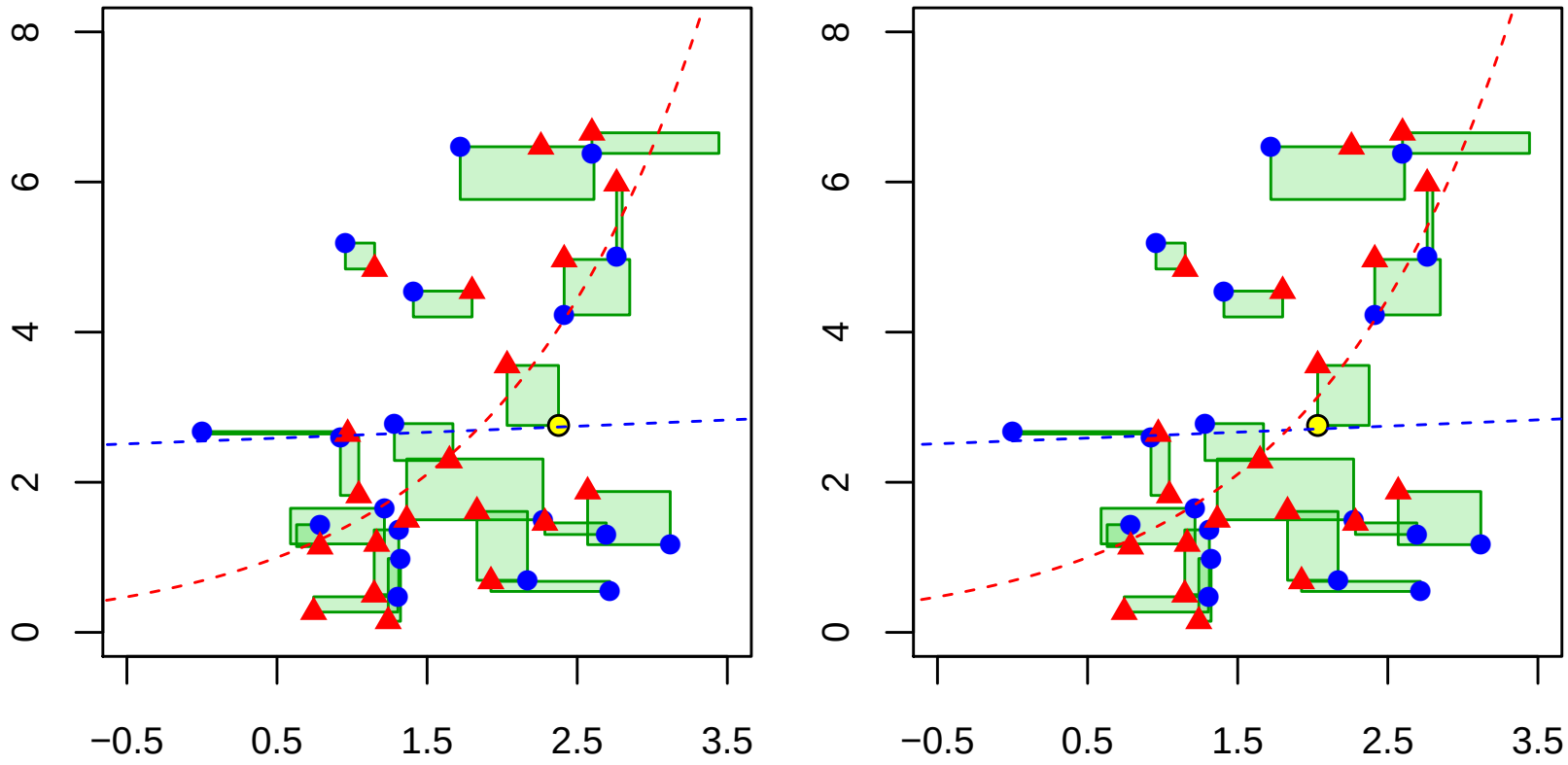


Abbildung: Unterschiede der Ergebnisse für die unabhängige Suche ausgehend von den Ecken (links) und der direkten Optimierung (rechts) für β_1 in einem generalisierten linearen Modell ohne Intercept.

Simulationsbeispiele: Spezialfall skalares x

- Für den Spezialfall mit skalarem x kann die Lösung analytisch bestimmt werden.
- Die direkte Optimierung des Parameters und die Suche ausgehend von den Ecken konnte nicht verwendet werden, da für L-BFGS-B bei Intervallbreiten von Null die Approximation des Gradienten fehlschlägt.
- Dies könnte aber durch Anpassung des Verfahrens ermöglicht werden (x als festen Parameter betrachten).
- Alle anderen Verfahren konnten aber stets die richtige Lösung bestimmen.

Zusammenfassung

- Verschiedene Ansätze zur Bestimmung des Identifizierungsbereichs wurden besprochen.
- Für Spezialfälle ist eine analytische oder zuverlässige Berechnung möglich.
- Die allgemeinen Verfahren beruhen entweder auf der numerischen Optimierung einer hochdimensionalen Zielfunktion oder heuristischer Suche.
- Es konnte kein Verfahren gefunden werden, bei dem man davon ausgehen kann, dass es im Allgemeinen die richtige Lösung liefert.
- Dennoch in den Beispielen immer gute Approximation oder exakte Bestimmung der Lösung.

Ausblick

- Weitere Simulationsbeispiele mit Vergleich der Methoden.
- Untersuchung anderer Verteilungen in generalisierten linearen Modellen.
- Modelle mit mehreren Kovariablen.
- Verwendung weiterer numerischer Optimierungsverfahren.
- Weitere Ideen für die heuristische Suche.
- Optimierung mit Gleichheitsnebenbedingung.
- Exakte Lösung für weitere Spezialfälle.
- Simultaner Identifizierungsbereich bei mehrdimensionalen Parametern.
- Konfidenzintervalle
  [Imbens und Manski, 2004, Chernozhukov et al., 2007, Stoye, 2009].
- Entwicklung eines R-Pakets.

Literatur I


Augustin, T. (2012).
Statistical analysis under data imprecision.
Unpublished Manuscript, Ludwig-Maximilians-Universität München.


Beresteanu, A. und Molinari, F. (2012).
Partial Identification Using Random Set Theory.
Journal of Econometrics *166*, 17–32.


Byrd, R. H., Lu, P., Nocedal, J. und Zhu, C. (1995).
A Limited Memory Algorithm for Bound Constrained Optimization.
SIAM Journal on Scientific Computing *16*, 1190–1208.


Chernozhukov, V., Hong, H. und Tamer, E. (2007).
Estimation and Confidence Regions for Parameter Sets in Econometric Models.
Econometrica *75*, 1243–1284.


Imbens, G. W. und Manski, C. F. (2004).
Confidence Intervals for Partially Identified Parameters.
Econometrica *72*, 1845–1857.


Lange, K. (2010).
Numerical Analysis for Statisticians.
Statistics and Computing, 2. Auflage, Springer.


Manski, C. F. (2003).
Partial Identification of Probability Distributions.
Springer, New York.

Literatur II


Moon, H. R. und Schorfheide, F. (2012).
Bayesian and Frequentist Inference in Partially Identified Models.
Econometrica *80*, 755–782.


Nelder, J. A. und Mead, R. (1965).
A Simplex Method for Function Minimization.
The Computer Journal *7*, 308–313.


Nocedal, J. (1980).
Updating Quasi-Newton Matrices with Limited Storage.
Mathematics of Computation *35*, 773–782.


R Development Core Team (2012).
optim: General-purpose Optimization.
Online: `http://www.r-project.org`.
R stats, version 2.15.1.


Rohwer, G. und Pötter, U. (2001).
Grundzüge der sozialwissenschaftlichen Statistik.
Grundlagentexte Soziologie, Beltz Juventa.


Stoye, J. (2009).
More on Confidence Intervals for Partially Identified Parameters.
Econometrica *77*, 1299–1315.