

4. Modellmittelung

In der Statistik werden Inferenz und Modellselektion meist als zwei separate Prozesse betrachtet: Wird aufgrund eines Kriteriums oder Verfahrens ein einziges, bestes Modell M_{κ^*} aus einer Menge von Kandidatenmodellen $\mathcal{M} = \{M_1, \dots, M_k\}$ ausgewählt, so wird dieses Modell implizit als korrekt angenommen und die dazugehörigen Parameterschätzungen und Konfidenzintervalle werden so bestimmt, als wäre das Modell *a priori* festgelegt worden. Dies ist äußerst problematisch, da der Selektionsschritt in der Regel datenbasiert ist, die Daten oft mehrere Modelle stützen und die Konstruktion eines korrekten Schätzers *post model selection*¹⁵ somit die Selektionsprozedur und die dadurch verursachte Unsicherheit bezüglich der Wahl von M_{κ^*} berücksichtigen sollte. Die Vernachlässigung dieses Aspekts kann ernste Folgen haben; insbesondere muss damit gerechnet werden, dass die Parameterschätzungen verzerrt sind, deren Varianz systematisch unterschätzt wird und die entsprechenden Konfidenzintervalle damit durchweg zu optimistisch sind, vergleiche hierzu auch Chatfield (1995), Draper (1995), Candolo, Davison und Demétrio (2003), Hjort und Claeskens (2003), Leeb und Pötscher (2005) sowie die dort angegebenen Referenzen.

Die Unsicherheit in der Modellwahl bezieht sich dabei meist auf zwei Aspekte: 1) die Wahl einer geeigneten, möglicherweise multivariaten Wahrscheinlichkeitsverteilung $f(\cdot)$ und 2) die Wahl einer dieser Verteilung zugehörigen Parametrisierung. Betrachtet man beispielsweise im Regressionskontext eine Zielgröße y , die aus Zähldaten besteht, sowie eine Menge potentieller Kovariablen X_1, \dots, X_p , so kann zur Modellierung der Abhängigkeitsstruktur $f(y|X_1, \dots, X_p)$ etwa ein 1) Poisson-, Negativ-Binomial- bzw. Quasi-Poisson-Ansatz verfolgt werden; die Wahl einer 2) entsprechend geeigneten Parametrisierung $f(y|X_1, \dots, X_p; \theta)$ ist in diesem Kontext äquivalent mit der Wahl geeigneter Regressoren. Unabhängig vom gewählten Verteilungsmodell führen verschiedene Kovariablenkombinationen und -transformationen zu einer unterschiedlichen Parametrisierung

¹⁵ Geht dem Inferenzprozess ein datenbasierter Modellselektionsschritt voraus, so wird der zugehörige Parameterschätzer in der Literatur häufig auch als *Post-Model-Selection-Estimator* (PMSE) bezeichnet.

und damit zu einer unterschiedlichen Interpretation der Datenstruktur. Insbesondere der zweite der eben genannten Aspekte, also die Unsicherheit bezüglich der Wahl einer geeigneten Parametrisierung anhand der Selektion von Variablen, ist Schwerpunkt dieses Kapitels.

Bereits seit Ende der 1970er Jahre wird die Problematik einer Vernachlässigung der Modellselektion in der Inferenz auch als solche erkannt (z.B. bei Leamer (1978), Hjort (1982), Hodges (1987), Pötscher (1991)); zu einer konkreten Ausarbeitung an Lösungsansätzen führten jedoch erst einige richtungsweisende Artikel Mitte der 1990er Jahre, im Speziellen die viel zitierten Arbeiten von Draper (1995) und Chatfield (1995), der hierzu bemerkt:

„In practice model uncertainty is a fact of life and likely to be more serious than other sources of uncertainty which have received far more attention from statisticians“

Er ergänzt in diesem Zusammenhang folgerichtig:

„[...] we must get over the key message that the properties of an estimator may depend not only on the selected model, but also on the selection process“

Eine Möglichkeit die Unsicherheit innerhalb der Modellselektion explizit zu berücksichtigen, besteht darin, sich bei der Inferenz nicht nur auf ein einziges Modell, sondern auf mehrere Modelle zu stützen. Das Kombinieren mehrerer Modelle ist Thema dieses Kapitels und wird auch als *Modellmittelung* bezeichnet. Die Literatur unterscheidet dabei zwischen bayesianischer Modellmittelung (Bayesian Model Averaging, BMA, Abschnitt 4.1) und frequentistischer Modellmittelung (Frequentist Model Averaging, FMA, Abschnitt 4.2).

Ungeachtet der zugrundeliegenden Philosophie beschäftigen sich die nächsten beiden Abschnitte mit der Konstruktion gewichteter Parameterschätzungen der Form

$$\hat{\theta} = \sum_{\kappa=1}^k w_{\kappa} \hat{\theta}_{\kappa}, \quad (4.1)$$

und deren Varianz $\text{Var}(\hat{\theta})$. Dabei beschreibt $\hat{\theta}_{\kappa}$ die Parameterschätzung im Modell $M_{\kappa} \in \mathcal{M} = \{M_1, \dots, M_k\}$, die gemäß ihrer Evidenz oder aufgrund anderer Überlegungen ein Gewicht w_{κ} , $\sum_{\kappa} w_{\kappa} = 1$, zugewiesen bekommt, wodurch explizit die Unsicherheit bezüglich der Modellwahl mitberücksichtigt wird. Das Hauptaugenmerk liegt dabei stets auf der Konstruktion der Gewichte, wie auch der Optimalität der daraus resultierenden gewichteten Parameterschätzung. Es ist zu beachten, dass für jede Vektorkomponente $\hat{\theta}_{\kappa i} \in \hat{\theta}_{\kappa}$ eines Modells M_{κ} , die nicht in einem konkurrierenden Modell M_{λ} , $\lambda \neq \kappa$, vertreten ist, stets angenommen wird, dass $\hat{\theta}_{\lambda i} = 0$.

Der Schätzer (4.1) wird auch als *BMA-Schätzer* bezeichnet wenn er bayesianisch motiviert ist bzw. als *FMA-Schätzer* wenn er frequentistisch motiviert ist. Er enthält als Spezialfall auch jeden Modellselektionsschätzer, da hierbei schlicht das Modell M_{κ^*} , das durch ein Verfahren oder Kriterium Γ gewählt wird, das Gewicht $w_{\kappa^*} = 1$ erhält, alle anderen Modelle das Gewicht 0. Diese Schätzer werden im Folgenden auch als *BMS-Schätzer* (Bayesian Model Selection) bzw. *FMS-Schätzer* (Frequentist Model Selection) bezeichnet.

4.1 Der bayesianische Ansatz

Die Idee bayesianischer Modellmittelung geht zurück auf Leamer (1978). Erst der rasante Fortschritt in der Entwicklung statistischer Software ermöglichte jedoch eine Implementierung und Weiterentwicklung dieser Ideen und resultierte in einer Vielzahl an Literatur innerhalb der letzten 15 Jahre. Eine ausführliche Beschreibung der Grundkonzepte sowie die Diskussion gängiger Probleme und Aspekte der Programmierung finden sich unter anderem bei Draper (1995), Kass und Raftery (1995), Chatfield (1995), Raftery, Madigan und Hoeting (1997) und Hoeting et al. (1999).

Der im bayesianischen Kontext intuitivste Gewichtungsansatz besteht darin, die Modelle $M_\kappa \in \mathcal{M} = \{M_1, \dots, M_k\}$ gemäß ihrer posteriori-Wahrscheinlichkeit

$$p(M_\kappa|y) = \frac{p(M_\kappa) \int_{\Theta_\kappa} p(y|M_\kappa, \theta_\kappa) \cdot p(\theta_\kappa|M_\kappa) d\theta_\kappa}{\sum_{\kappa=1}^k p(M_\kappa) \int_{\Theta_\kappa} p(y|M_\kappa, \theta_\kappa) \cdot p(\theta_\kappa|M_\kappa) d\theta_\kappa}, \quad (4.2)$$

wie in (3.33) beschrieben, zu gewichten, also für (4.1) die Gewichte

$$w_\kappa^{(1)} = p(M_\kappa|y) \quad (4.3)$$

zu wählen. Aus (3.34) und (3.35) folgt, dass $p(M_\kappa|y) \approx -\frac{1}{2}\text{SBC}$ und somit ergibt sich als eine weitere Möglichkeit der Gewichtung

$$w_\kappa^{(2)} = \frac{\exp(-\frac{1}{2}\text{SBC}_\kappa)}{\sum_{\kappa=1}^k \exp(-\frac{1}{2}\text{SBC}_\kappa)}. \quad (4.4)$$

Bezeichne $\hat{\theta}_\kappa$ einen bayesianischen Punktschätzer im Modell M_κ , beispielsweise den Erwartungswert oder den Modus der posteriori-Verteilung $p(\theta_\kappa|M_\kappa, y)$, dann lässt sich unter Verwendung von (4.3) oder (4.4) ein gewichteter Parameterschätzer $\hat{\hat{\theta}}$ gemäß (4.1) bestimmen. Seine Varianz erhält man über

$$\text{Var}(\hat{\theta}) = \mathbb{E}_{\mathcal{M}}(\text{Var}(\hat{\theta}_{\kappa}|y, M_{\kappa})) + \text{Var}_{\mathcal{M}}(\mathbb{E}(\hat{\theta}_{\kappa}|y, M_{\kappa})), \quad (4.5)$$

vergleiche auch Draper (1995). Der erste Term repräsentiert dabei die gewichtete Varianz innerhalb der Modelle, der zweite Term die gewichtete Varianz zwischen den Modellen, wodurch die Unsicherheit bezüglich der Modellselektion implizit berücksichtigt wird.

Der bayesianische Ansatz unterliegt einigen generellen, vorwiegend technischen Beschränkungen, deren Relevanz im Folgenden diskutiert werden soll:

- Ist die Anzahl der Kandidatenmodelle sehr groß, so erweist sich die Umsetzung der Grundkonzepte als äußerst schwierig bzw. unmöglich. Madigan und Raftery (1994) schlagen daher eine Reduzierung der in die Summation aufzunehmenden Modelle vor: Es sollen nur die Modelle verwendet werden, die – gemessen an ihrer posteriori-Wahrscheinlichkeit – eine ausreichende Erklärungskraft besitzen, also die Menge

$$\mathcal{A}' = \left\{ M_{\kappa} : \frac{\max_{M_{\kappa} \in \mathcal{M}} \{p(M_{\kappa}|y)\}}{p(M_{\kappa}|y)} \leq C \right\},$$

wobei C eine Konstante beschreibt. Madigan und Raftery (1994) sowie Raftery, Madigan und Volinsky (1996) verwenden bei der Auswertung von Daten dabei einen Wert von $C = 20$. In Anlehnung an Occam's Razor (vgl. Kapitel 2) sollen sparsame Modelle bevorzugt, komplexere Modelle ausgeschlossen werden. Modelle der Menge

$$\mathcal{B} = \left\{ M_{\kappa} : \exists M_{\lambda} \in \mathcal{A}', M_{\lambda} \subset M_{\kappa}, \frac{p(M_{\lambda}|y)}{p(M_{\kappa}|y)} > 1 \right\}$$

werden daher nicht weiter betrachtet. Gemäß Madigan und Raftery (1994) umfasst eine statistische Analyse somit die oben vorgestellten bayesianischen Modellmittelungsmethoden, angewandt auf die Menge $\mathcal{A} = \mathcal{A}' \setminus \mathcal{B} \subseteq \mathcal{M}$. Dieses Vorgehen wird auch als *Occam's Window* bezeichnet. Hinweise zur Implementierung finden sich bei Volinsky et al. (1997) und Hoeting et al. (1999).

- Die in (4.2) enthaltenen Integrale sind analytisch nicht immer zu bestimmen. Eine Approximation liefert jedoch meist die Laplace-Methode (vgl. Tierney und Kadane (1986)). In vielen Fällen kann auch die Approximation durch das Schwarzsche

Bayes-Kriterium verwendet werden, vergleiche (4.4). Dies ermöglicht eine einfache Implementierung und ist Grundlage des Pakets „BMA“ (Bayesian Model Averaging, Raftery et al. (2006)) für die statistische Software *R*. Wie in Abschnitt 3.4.1 detailliert beschrieben, ist diese Approximation unter Umständen jedoch äußerst ungenau. Insofern ist ein solches Vorgehen mit Vorsicht zu genießen.

- Die Bestimmung der priori-Wahrscheinlichkeiten $p(M_\kappa)$ ist unklar. In einem Großteil der Literatur sowie in zahlreichen Anwendungen der bayesianischen Modellmittelung wird schlicht auf die Spezifikation der Wahrscheinlichkeiten verzichtet, indem alle Modelle a priori als gleich wahrscheinlich eingestuft werden, also $p(M_\kappa) = \frac{1}{k}$, $\forall M_\kappa \in \mathcal{M}$, gewählt wird. Meist wird sogar die SBC-Approximation (4.4) verwendet, wodurch implizit keine priori-Information für die Modelle spezifiziert werden muss. Diese Vorgehensweise erlaubt eine einfache und zielgerichtete Implementierung. Das bayesianische Grundverständnis priori-Wissen zu nutzen geht dabei jedoch vollständig verloren.

Nur sehr wenige Arbeiten beschäftigen sich mit der Frage, was die priori-Wahrscheinlichkeiten $p(M_\kappa)$ wirklich bedeuten, wie sie zu interpretieren sind und wie sie konstruiert werden können. Hoeting et al. (1999) greifen eine Idee von George und McCulloch (1993) auf, und schlagen vor, für parametrische Regressionsmodelle die priori-Wahrscheinlichkeiten wie folgt zu wählen:

$$p(M_\kappa) = \prod_{j=1}^p \pi_j^{\delta_{\kappa j}} (1 - \pi_j)^{1 - \delta_{\kappa j}}.$$

Dabei beschreibt $\pi_j \in [0, 1]$ die priori-Wahrscheinlichkeit für $\beta_j \neq 0$ im Regressionsmodell M_κ . Die Indikatorvariable $\delta_{\kappa j}$ gibt an, ob sich die Variable j im Modell M_κ befindet oder nicht. Dieser Ansatz ist sicherlich eine erste Möglichkeit Modelle als mehr oder weniger wahrscheinlich einzustufen. Problematisch erscheint jedoch die implizite Annahme, dass das Vorhandensein oder Fehlen der einzelnen Variablen voneinander unabhängig ist.

4.2 Frequentistische Ansätze

Der hohe computationale Aufwand zur Berechnung der posteriori-Wahrscheinlichkeiten (4.2), die daraus resultierende lange Rechenzeit, die unbefriedigende Approximation

über das SBC-Kriterium sowie die unklare Bestimmung und Interpretation der prior-Wahrscheinlichkeiten $p(M_\kappa)$ führten seit Ende der 1990er Jahre zu einer Diskussion über frequentistische Alternativen in der Modellmittelung und zur Veröffentlichung einiger Artikel, die konkrete Lösungsansätze für die Konstruktion eines FMA-Schätzers entwickeln (Buckland, Burnham und Anderson (1997), Hansen (2007), Liang et al. (2010)) oder ein einheitliches frequentistisches Paradigma entwerfen (Hjort und Claeskens (2003)). In den folgenden Abschnitten 4.2.1-4.2.3 werden diese Ansätze vorgestellt und diskutiert. Abschnitt 4.2.4 erläutert wie in diesem Zusammenhang Varianzschätzungen konzipiert werden können und Abschnitt 4.2.5 überträgt einige Kernkonzepte in den Kontext der Faktorenanalyse. Weitere Beiträge und Diskussionen zu dieser Thematik finden sich auch in den hier nicht näher kommentierten Arbeiten von Yang (2001, 2003), Yuan und Yang (2005), Hjort und Claeskens (2006), Leung und Barron (2006), Hansen (2008a,b, 2009) sowie Hansen und Racine (2009).

4.2.1 Kriteriumsasierte Schätzungen

Eine Möglichkeit die Evidenz eines Modells zu beurteilen, liegt in der Betrachtung eines beliebigen Selektionskriteriums Γ wie in Kapitel 3 beschrieben. Diese simple Einsicht kann dazu verwendet werden, Gewichte für die Kandidatenmodelle zu konstruieren. Der in diesem Zusammenhang erste in der Literatur beachtete, nicht-bayesianische Ansatz zur Modellmittelung geht auf Buckland, Burnham und Anderson (1997) zurück. Die Autoren schlagen vor, die Gewichte¹⁶

$$w_\kappa^{(3)} = \frac{\exp(-\frac{1}{2}AIC_\kappa)}{\sum_{\kappa=1}^k \exp(-\frac{1}{2}AIC_\kappa)} \quad (4.6)$$

zu verwenden und damit Modelle mit einem geringeren AIC-Wert als plausibler einzustufen und folglich auch höher zu gewichten. Im Prinzip kann anstelle des Kriteriums von Akaike auch jedes andere sinnhafte Kriterium Γ verwendet werden, siehe hierzu auch Abschnitt 5.2.2. Dieser Ansatz erscheint insgesamt plausibel, auch unter Berücksichtigung von (4.4). Dennoch ist die Konstruktion ad-hoc; die Arbeiten von Hjort und Claeskens (2003), Hansen (2007) sowie Liang et al. (2010) zeigen ein in Simulationsstudien bezüglich dem MSE durchweg schlechtes Abschneiden der zugehörigen FMA-Schätzung $\hat{\hat{\theta}} = \sum_{\kappa=1}^k w_\kappa^{(3)} \hat{\theta}_\kappa$. Aufgrund der einfachen Implementierung hat das Konzept jedoch

¹⁶ Die Gewichte $w_\kappa^{(3)}$ werden in der Literatur auch häufig als *Akaike Gewichte*, *AIC-Gewichte* oder *geglättete AIC-Gewichte* bezeichnet.

bereits eine breite Verwendung gefunden, insbesondere in Biologie, Ökonometrie und Medizin. Als Beispiele für eine erfolgreiche und anschauliche Umsetzung sind unter anderem die Arbeiten von Hayes, Weikel und Huso (2003), Reid et al. (2003), Candolo, Davison und Demétrio (2003) und Mackenzie et al. (2005) zu nennen.

Buckland, Burnham und Anderson (1997) beschreiben ferner die Möglichkeit, Gewichte über Bootstrapping zu erhalten. Sei B die Anzahl der Bootstrap-Stichproben und b_κ die Anzahl der Fälle in denen das Modell $M_\kappa \in \mathcal{M} = \{M_1, \dots, M_k\}$ aufgrund von Akaikes Informationskriterium (oder eines anderen Verfahrens Γ) gewählt wird. Dann ergeben sich als mögliche Gewichte

$$w_\kappa^{(4)} = \frac{b_\kappa}{B}. \quad (4.7)$$

Burnham und Anderson (2002, Seite 90 ff.) bemerken, dass für verlässliche Schätzungen meist 10.000 Bootstrap-Stichproben oder mehr gezogen werden sollten. Aufgrund des hohen computationalen Aufwands und der großen Menge an alternativen Gewichtsschätzungen wird diese Methodik eher selten verwendet.

4.2.2 Der MMA-Schätzer

Gegeben sei eine Menge an Regressoren $\mathcal{X} = \{X_1, \dots, X_k\}$ sowie eine Menge an linearen Modellen $\mathcal{M} = \{M_1, \dots, M_k\}$, so dass jedes Modell $M_\kappa \in \mathcal{M}$, $\kappa = 1, \dots, k$, die Gestalt

$$y = \sum_{j=1}^{\kappa} \beta_j X_j + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I),$$

besitzt. Dies bedeutet implizit, dass die k Regressoren geordnet vorliegen müssen und die Menge \mathcal{M} der Kandidatenmodelle aus k verschachtelten Modellen besteht. Ein gewichteter Parameterschätzer zur Modellmittelung ist für diese Situation gemäß (4.1) durch $\hat{\hat{\beta}} = \sum_{\kappa=1}^k w_\kappa \hat{\beta}_\kappa$ gegeben, wobei $\hat{\beta}_\kappa$ den Parameterschätzer aus Modell M_κ bezeichnet. Daraus ergibt sich der geschätzte, gewichtete Erwartungswert zu $\hat{\mu}_w = X_k \hat{\hat{\beta}}$. Hansen (2007) definiert in Anlehnung an Mallows C_p (Abschnitt 3.2.1) das Kriterium

$$\tilde{C}_p = (y - X_k \hat{\hat{\beta}})'(y - X_k \hat{\hat{\beta}}) + 2\sigma^2 K_w, \quad (4.8)$$

wobei $K_w = \text{sp}(P_w)$ mit $P_w = \sum_{\kappa=1}^k w_{\kappa} P_{\kappa}$ ¹⁷, $P_{\kappa} = X_{\kappa}(X'_{\kappa}X_{\kappa})^{-1}X'_{\kappa}$. Die Varianz σ^2 kann dabei über die KQ-Schätzung im vollen Modell M_k geschätzt werden. Er schlägt vor, die Gewichte zur Modellmittelung so zu wählen, dass

$$w_{\kappa}^{(5)} = \arg \min_{w_{\kappa} \in \mathcal{H}} \tilde{C}_p, \quad (4.9)$$

wobei $\mathcal{H} = \{w_{\kappa} \in [0, 1]^k : \sum_{\kappa=1}^k w_{\kappa} = 1\}$. Der zugehörige FMA-Schätzer $\hat{\beta} = \sum_{\kappa=1}^k w_{\kappa}^{(5)} \hat{\beta}_{\kappa}$ wird in Anlehnung an Mallows C_p als Mallows-Model-Averaging-Schätzer, kurz *MMA-Schätzer*, bezeichnet. Da der erste Term in (4.8) offensichtlich quadratisch bezüglich w_{κ} ist, der zweite linear, lassen sich die Gewichte über quadratische Optimierung ermitteln. Statistische und mathematische Software beinhalten in der Regel passende numerische Algorithmen, so dass die Berechnung von (4.9) prinzipiell unproblematisch ist.

Der Ansatz von Hansen ist der erste, der Gewichte unter dem Gesichtspunkt der Optimalität konstruiert. Konkret sind damit die folgenden beiden Charakteristika gemeint:

- (i) Die Verwendung der gewichteten Schätzung $\hat{\beta} = \sum_{\kappa=1}^k w_{\kappa}^{(5)} \hat{\beta}_{\kappa}$ und damit von $\hat{\mu}_w = X_k \hat{\beta}$ garantiert unter bestimmten Voraussetzungen ein asymptotisch effizientes Verfahren ähnlich (3.53)¹⁸, vergleiche Hansen (2007, Theorem 1).
- (ii) $E(\tilde{C}_p) = E[(\hat{\mu}_w - \mu)'(\hat{\mu}_w - \mu)] + n\sigma^2$, vergleiche Hansen (2007, Lemma 3). In Analogie zu Mallows C_p steht damit die Minimierung des mittleren quadratischen Vorhersagefehlers im Vordergrund. Der einzige Unterschied liegt in der Betrachtung des Risikos für den gemittelten Erwartungswert $\hat{\mu}_w$ anstelle des Erwartungswerts auf Basis von (3.4).

Gleichwohl ergeben sich Probleme, die eine praktische Umsetzung des MMA-Schätzers schwierig gestaltet: Eine natürliche Anordnung der Regressoren existiert fast nie und eine willkürliche Anordnung kann das Ergebnis entscheidend beeinflussen, vergleiche hierzu

¹⁷ P_w kann als gewichtete Projektionsmatrix interpretiert werden. Diese ist zwar symmetrisch, i.d.R. jedoch nicht mehr idempotent.

¹⁸ Es ist für (3.53) lediglich erforderlich, den gemittelten Erwartungswert $\hat{\mu}_w$ an Stelle von $\mu(M_{\kappa})$ zu betrachten und die Optimalität für alle $w_{\kappa} \in \mathcal{H}$ anstatt für alle $M_{\kappa} \in \mathcal{M}$ zu fordern. Hansen (2007) zeigt damit, dass $L(\mathcal{M}; w^{(5)}) / \inf_{w \in \mathcal{H}} L(\mathcal{M}; w) \xrightarrow{p} 1$, wobei $L(\mathcal{M}; w) = \|\mu - \hat{\mu}_w\|^2 / n$. Sein Nachweis unterliegt insofern recht strengen Voraussetzungen, als dass gefordert wird, dass die Gewichte der diskreten Menge $\mathcal{H}_N = \{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$, $N \in \mathbb{N}$, entstammen. Wan, Zhang und Zou (2010) zeigen jedoch später, dass diese Einschränkung nicht nötig ist und der Nachweis der asymptotischen Effizienz für jede beliebige Menge \mathcal{H} gilt.

auch Liang et al. (2010). Zudem beschränkt sich die Verwendung auf das lineare Modell, was ein entscheidender Nachteil gegenüber allen bisher vorgestellten bayesianischen wie auch frequentistischen Ansätzen darstellt. Eine ausführlichere Diskussion dieser und anderer Aspekte erfolgt in den Simulationsstudien in Kapitel 6.

4.2.3 Der OPT-Schätzer

Der OPT-Schätzer von Liang et al. (2010) ist – motiviert durch die Arbeit von Hansen (2007) – ebenfalls unter den Gesichtspunkten der Optimalität konstruiert, lehnt sich in Notation und Denkweise jedoch stark an die Pretest-Literatur, insbesondere die Artikel von Magnus und Durbin (1999), Danilov und Magnus (2004) und Magnus, Powell und Prüfer (2008) an. Gegeben sei ein lineares Regressionsmodell $y = X\alpha + Z\gamma + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$, bei dem X eine $n \times d$ Designmatrix beschreibt, die diejenigen Regressoren enthält, die aus theoretischen oder anderen Gründen auf jeden Fall in ein Endmodell aufgenommen werden sollten und Z eine $n \times q$ Designmatrix, die diejenigen Regressoren enthält, die potentiell in ein Endmodell aufgenommen werden können. Der $d \times 1$ Vektor α ist dabei der Parameter von Interesse, der $q \times 1$ Vektor γ wird als nuisance Parameter behandelt, um eine „bessere“ Schätzung von α zu erhalten; insgesamt können also 2^q verschiedene Kandidatenmodelle betrachtet werden. Der restringierte ($\gamma = 0$) und der unrestringierte KQ-Schätzer von α ergibt sich jeweils zu $\hat{\alpha}_r = (X'X)^{-1}X'y$ bzw. $\hat{\alpha}_u = \hat{\alpha}_r - Q\hat{\zeta}$, wobei $Q = (X'X)^{-1}X'ZC$, $C = (Z'\{I - P\}Z)^{-\frac{1}{2}}$, $P = X(X'X)^{-1}X'$, $\hat{\zeta} = CZ'(I - P)y$, $\hat{\zeta} \sim N(\zeta, \sigma^2 I_q)$, vergleiche auch Magnus und Durbin (1999). Betrachtet man alle 2^q Kandidatenmodelle, die sich aus den q potentiellen Einflussfaktoren $Z = (Z_1, \dots, Z_q)$ ergeben, so definiert sich der κ -te, $\kappa = 1, \dots, 2^q$, restringierte KQ-Schätzer $\hat{\alpha}_\kappa$ von α als

$$\hat{\alpha}_\kappa = \hat{\alpha}_r - QW_\kappa \hat{\zeta}$$

mit $W_\kappa = I_q - P_\kappa$, $P_\kappa = CS_\kappa(S'_\kappa C^2 S_\kappa)^{-1}S'_\kappa C$. Dabei bezeichnet S'_κ eine $r \times q$ Selektionsmatrix, $r = \dim(M_\kappa)$, die gemäß der Restriktion $S'_\kappa \gamma = 0$ angibt, welche Kovariablen $Z^* \subseteq Z$ im κ -ten Modell von Interesse sind. Für den nuisance Parameter γ erhält man entsprechend die KQ-Schätzung $\hat{\gamma}_\kappa = CW_\kappa \hat{\zeta}$. Der *OPT-Schätzer* ist derjenige FMA-Schätzer

$$\hat{\hat{\alpha}} = \sum_{\kappa=1}^{2^q} w_\kappa^{(6)} \hat{\alpha}_\kappa, \quad (4.10)$$

der gemäß dem schwachen MSE-Kriterium (Wallace (1972)) die Gewichte $w_\kappa^{(6)}$ so wählt, dass

$$w_\kappa^{(6)} = \arg \min_{w_\kappa \in \mathcal{H}} \text{sp}\{\widehat{MSE}(\hat{\alpha})\}, \quad (4.11)$$

wobei $\mathcal{H} = \{w_\kappa \in [0, 1]^{2^q} : \sum_{\kappa=1}^{2^q} w_\kappa = 1\}$. Auf Basis dieser Gewichte erhält man entsprechend den gewichteten KQ-Schätzer $\hat{\gamma} = \sum_{\kappa=1}^{2^q} w_\kappa^{(6)} \hat{\gamma}_\kappa$ für den nuisance Parameter γ . Ein alternativer OPT-Schätzer, der unter Aspekten der Vorhersagequalität verwendet werden kann, entspricht gemäß Liang et al. (2010) demjenigen FMA-Schätzer, der die Gewichte (4.11) verwendet, die nicht die Spur des geschätzten MSE von $\hat{\alpha}$ minimieren, sondern die Spur des MSE von $\hat{\mu} = H \sum_{\kappa=1}^{2^q} w_\kappa^{(7)} \hat{\beta}_\kappa$, $\hat{\beta}_\kappa = (\hat{\alpha}'_\kappa, \hat{\gamma}'_\kappa)'$, $H = (X, Z)$.

Der OPT-Schätzer stellt gegenüber dem MMA-Schätzer insofern eine klare Weiterentwicklung dar, als dass auch ohne eine explizite Anordnung der Regressoren die geforderte Optimalitätseigenschaft (nämlich die Gültigkeit von (4.11)) erhalten bleibt; die Simulationen von Liang et al. (2010) unterstützen diese Annahme ausnahmslos und zeigen, dass willkürliche Umordnungen der Regressoren einen entscheidenden Einfluss auf das Verhalten des MMA-Schätzers besitzen.

Problematisch erscheint jedoch der dem Ansatz innewohnende, hohe computationale Aufwand. Zum einen erfordert die explizite Minimierung der Spur des MSE von $\hat{\alpha}$ einen großen technischen Einsatz, vergleiche hierzu auch Liang et al. (2010, Theorem 2); zum anderen ergibt sich aus der Notwendigkeit über alle 2^q Modelle zu mitteln, bei einer großen Anzahl an potentiellen Einflussgrößen ein Rechenaufwand, der für eine explizite Berechnung des OPT-Schätzers zu hoch sein kann.

4.2.4 Schätzung der Varianz

Um Konfidenzintervalle für einen FMA-Schätzer bestimmen zu können, wird eine Schätzung für dessen Varianz benötigt. Derzeit werden hierfür vor allem zwei Ansätze diskutiert und verwendet: Der erste findet seinen Ursprung in der Arbeit von Buckland, Burnham und Anderson (1997), die vorschlagen die Varianz von $\hat{\theta}$ gemäß

$$\widehat{\text{Var}}(\hat{\theta}) = \left\{ \sum_{\kappa=1}^k w_\kappa \sqrt{\widehat{\text{Var}}(\hat{\theta}_\kappa | M_\kappa) + (\hat{\theta}_\kappa - \hat{\theta})^2} \right\}^2 \quad (4.12)$$

Generalized Additive Model based Recursive Imputation

- Schritt 1:** Sei s ganzzahlig und beginne mit $s = 1$.
- Schritt 2:** Bezeichne D_s den Subdatensatz, der nur die Zeilen mit s fehlenden Werten enthält.
- Schritt 3:** \mathcal{X}^* ist die Menge aller Variablen, die fehlende Werte enthalten. Wähle $X_j \in \mathcal{X}^*$ mit der geringsten Anzahl an fehlenden Werten in D_s .
- Schritt 4:** Für jeden fehlenden Wert $x_{ij} \in X_j$, $i = 1, \dots, n$, und $x_{ij} \in D_s$,
- a) fitte das GAM,

$$\eta(X_j) = \alpha + f_0(y) + \sum_{\substack{l=1 \\ l \neq j, l \notin \Phi_l}}^p f_l(X_l) + \epsilon, \quad (5.13)$$

unter Verwendung der vollständigen Daten D_*^c , wobei $\Phi_l = \{l : x_{il} \text{ fehlend}\}$ und $\eta(\cdot)$ eine passende Linkfunktion ist. Wenn x_{ij} und x_{kj} gleichzeitig fehlen und $i < k$, dann erfolgt die Imputation von x_{ij} vor x_{kj} .

- b) imputiere einen neuen Wert \tilde{x}_{ij} für x_{ij} basierend auf dem geschätzten GAM

$$\eta(\tilde{x}_{ij}) = \hat{\alpha} + \hat{f}_0(y_i) + \sum_{\substack{l=1 \\ l \neq j, l \notin \Phi_l}}^p \hat{f}_l(x_{il}), \quad (5.14)$$

wobei $\hat{\alpha}$, \hat{f}_0 und \hat{f}_l Schätzungen von α , f_0 und f_l sind.

- c) Aktualisiere D_s , D_*^c und \mathcal{X}^* indem die imputierten Werte als vollständige Beobachtungen gewertet werden.

- Schritt 5:** Wiederhole ab Schritt 3 bis $D_s = \emptyset$.
- Schritt 6:** Setze s auf $s + 1$.
- Schritt 7:** Wiederhole ab Schritt 2 bis s die maximale Anzahl an fehlenden Werten in einer Zeile erreicht hat.
- Schritt 8:** Wiederhole die Schritte 1–7 bis $\mathcal{X}^* = \emptyset$.
-

Tab. 5.1: Algorithmus für eine verallgemeinerte Regressionsimputation

Für eine detaillierte Illustration des Algorithmus wird im Folgenden ein Beispieldatensatz (Tabelle 5.2) verwendet. Die Variablen y , X_1 und X_2 sind dabei metrisch, X_3 ist eine Zählvariable und X_4 binär. Insgesamt fehlen dem Datensatz fünf Werte: x_{41} , x_{23} , x_{24} , x_{34} und x_{54} .

i	y	X_1	X_2	X_3	X_4
1	1.2	24.2	-8.2	7	0
2	2.3	23.4	-3.2	*	*
3	2.2	30.0	-0.1	9	*
4	2.3	*	-4.1	10	1
5	2.6	22.0	-0.8	7	*
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
100	1.7	26.4	-2.9	7	1

Tab. 5.2: Beispieldaten zur Veranschaulichung des Algorithmus

Gemäß Schritt 1 und 2 wird $s = 1$ gesetzt und folglich all diejenigen Beobachtungen betrachtet, die exakt einen fehlenden Wert aufweisen; in Tabelle 5.2 sind dies genau die dritte, vierte und fünfte Zeile. Da X_1 weniger fehlende Werte besitzt als X_4 (Schritt 3) wird das generalisierte additive Modell $\eta(X_1) = \alpha + f_0(y) + f_2(X_2) + f_3(X_3) + f_4(X_4) + \epsilon$ auf Basis der vollständigen Fälle D_*^c gefittet um – unter Verwendung der Gauß’schen Linkfunktion – eine Imputation für x_{41} wie in Schritt 4 b) beschrieben zu erhalten. Diese Imputation kann dabei der konkrete, eventuell gerundete, Vorhersagewert \tilde{x}_{ij} aus (5.14) sein, oder aus der zugehörigen geschätzten Verteilung (hier: Normalverteilung) gezogen werden. Die vollständigen Fälle werden gemäß 4 c) aktualisiert und enthalten nun den imputierten Wert \tilde{x}_{41} . Schritt 5 gibt vor, im Weiteren den Wert x_{34} zu imputieren, anschließend x_{54} . Um dem Sachverhalt Rechnung zu tragen, dass es sich bei X_4 um eine binäre Variable handelt, betrachtet man das GAM $\eta(X_4) = \alpha + f_0(y) + f_1(X_1) + f_2(X_2) + f_3(X_3) + \epsilon$ mit logit-Link; die Imputation \tilde{x}_{34} für x_{34} erhält man weiterhin als Vorhersage dieser Regression. Analog erfolgt die Imputation von x_{54} . Da x_{41} , x_{34} und x_{54} imputiert worden sind, reduziert sich D_1 auf die leere Menge, $D_s = D_1 = \emptyset$. Daher wird infolge von Schritt 6 s auf $s = 2$ gesetzt und damit all diejenigen Beobachtungen herangezogen, die exakt zwei fehlende Werte aufweisen. Dies betrifft im vorliegenden Beispiel ausschließlich die zweite Zeile. Um einen imputierten Wert für x_{23} zu erhalten, wird – unter Verwendung der log-Linkfunktion aufgrund der Zählvariablen – das Modell $\eta(X_3) = \alpha + f_0(y) + f_1(X_1) + f_2(X_2) + \epsilon$ angepasst. Die Variable X_4 wird dabei nicht

verwendet, da x_{24} fehlt. Aus der Vorhersage dieses Modells (dies ist erneut entweder der exakte vorausgesagte Wert \tilde{x}_{23} oder ein zufällig gezogener Wert aus der entsprechenden $Po(\tilde{x}_{23})$ -Verteilung) wird erneut eine Imputation gewonnen. Die ständige Aktualisierung von D_s führt an dieser Stelle dazu, dass D_2 leer ist, da der einzige fehlende Wert, x_{24} , nun Teil der Menge D_1 ist. Da s die maximale Anzahl an fehlenden Werten in einer Zeile ($s = 2$) erreicht hat, wird s erneut auf $s = 1$ gesetzt (Schritt 7 und 8) und ein generalisiertes additives Modell mit y , X_1 , X_2 , X_3 als Kovariablen und der logit-Linkfunktion verwendet, um die letzte Imputation zu erhalten. Der Datensatz ist damit vollständig aufgefüllt.

Die Verwendung des Algorithmus kann als verteilungsbasierte Imputation interpretiert werden, da die GAM-basierte Imputation implizit die bedingte prädiktive Verteilung der fehlenden Werte $p(D^{\text{mis}}|D^{\text{obs}}; \theta)$ modelliert. Hierbei wird $\theta = (\alpha, \gamma)$ als fest angenommen und daher die Unsicherheit bei der Schätzung der Regressionsparameter nicht berücksichtigt. Diesem Umstand kann auf verschiedene Art und Weise Rechnung getragen werden: Gemäß (5.1) kann unter Berücksichtigung der – schwer zu bestimmenden – posteriori-Verteilung $p(\theta|D^{\text{obs}})$ die prädiktive a-posteriori-Verteilung $p(D^{\text{mis}}|D^{\text{obs}})$ bestimmt werden. Daraus lassen sich prinzipiell (multiple) Imputationen gewinnen. Praktisch bedeutet dies, dass nicht explizit die geschätzten Regressionsparameter sowie die geschätzte Varianz zur Imputation verwendet werden, sondern zufällige Ziehungen aus deren (posteriori-) Verteilung. Auch kann dem Regressionsmodell aus Schritt 4 b) ein stochastischer Fehlerterm hinzugefügt werden. Einige Beispiele wie sich dies realisieren lässt, finden sich bei Schafer (1997). Generell ist ein solches Vorgehen bei einer Erweiterung der Regressionsimputation auf additive Modelle schwer zu realisieren. Werden in dem vorgestellten Algorithmus jedoch anstelle generalisierter additiver Modelle gewöhnliche GLMs verwendet, so ist eine Umsetzung möglich. In Anlehnung an Schomaker, Wan und Heumann (2010) wird der oben vorgestellte Algorithmus als *Generalized Additive Model based Recursive Imputation* (GAMRI) bezeichnet, derselbe Algorithmus, der nur generalisierte lineare Regressionsmodelle verwendet, als *Generalized Linear Model based Recursive Imputation* (GLMRI).

Die Reihenfolge in der die Werte imputiert werden, die Wahl ob und wann die vollständigen Daten D_*^c aktualisiert werden und auf welche Art und Weise dem Imputationsmodell stochastische Komponenten hinzugefügt werden eröffnet viele Möglichkeiten der Konstruktion alternativer Algorithmen im Stil von GAMRI. Jede Wahl ist in gewisser Weise als ad-hoc anzusehen; dennoch liegt diesem Prozedere ein Grundprinzip zugrunde: die Verwendung der Abhängigkeitsstruktur der Variablen eines Datensatzes, um Impu-

tationen zu gewinnen und die für die Modellselektion und Modellmittelung interessanten Effekte zu erhalten.

Nächste Nachbarn Imputation

Die Kernidee einer Nächste-Nachbarn-Imputation besteht darin, die fehlenden Werte der Datenmatrix $x_{ij} \in D_*$ durch beobachtete Werte einer anderen, „ähnlichen“ Untersuchungseinheit $x_{lj} \in D_*$ zu ersetzen. Dieses sehr allgemeine Prinzip wird in der Literatur aus den unterschiedlichsten Blickwinkeln betrachtet, analytische Betrachtungen sind jedoch häufig restringiert auf den Spezialfall zweier stetiger Variablen von denen die eine vollständig beobachtet wurde und die andere nicht, vergleiche in etwa Chen und Shao (2000) sowie Nittner (2003). Generell ist allen Ansätzen jedoch gemein, dass sie für eine gegebene Metrik $d(\cdot, \cdot)$ den Abstand zwischen der Beobachtung mit dem fehlenden Wert und allen anderen Beobachtungen die keinen fehlenden Wert aufweisen berechnen und denjenigen Wert x_{lj} imputieren, der der Beobachtung $D_l = (y_l, \mathbf{x}_l)$ angehört, die den Abstand zu $D_i = (y_i, \mathbf{x}_i)$ auf Basis von $d(\cdot, \cdot)$ minimiert, $i \neq l$; diese Beobachtung wird auch *nächster Nachbar* genannt. Die Metrik ist dabei häufig die euklidische Distanz, angewandt auf die Beobachtungsvektoren $D_{i/j} = (y_i, x_{i1}, \dots, x_{i(j-1)}, x_{i(j+1)}, \dots, x_{ip})$ und $D_{l/j} = (y_l, x_{l1}, \dots, x_{l(j-1)}, x_{l(j+1)}, \dots, x_{lp})$ des zugehörigen Subdatensatzes $D_{/j} = \{y, X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p\}$. Alternativ können auch die k nächsten Nachbarn für die Imputation verwendet werden: Hierbei werden die k Beobachtungen betrachtet, die bezüglich der Metrik $d(\cdot, \cdot)$ den geringsten Abstand zur Beobachtung $D_{i/j}$ haben und der (gewichtete) Mittelwert dieser k Werte wird dann zur Imputation verwendet. Formell bedeutet dies, dass für einen fehlenden Wert $x_{ij} \in D_*$ die Menge der k nächsten Nachbarn,

$$D^{\text{kNN}} = \{x_{lj} | l = \arg \min_{l \neq i, l \notin \Phi_l} d(D_{i/j}, D_{l/j})\},$$

betrachtet wird, wobei kmin eine Funktion bezeichnet, die die k kleinsten ihrer Argumente zurückgibt und $\Phi_l = \{l : D_{l/j} \text{ enthält fehlende Werte}\}$. Das arithmetische Mittel aller $D_l^{\text{kNN}} \in D^{\text{kNN}}$, $l = 1, \dots, k$, liefert den Imputationswert

$$\tilde{x}_{ij} = \frac{1}{k} \sum_{l=1}^k D_l^{\text{kNN}}$$

für x_{ij} . Dieses Prinzip ist in dem R-Paket „EMV“ (Estimation of Missing Values, Gottardo (2008)) implementiert und wird im Folgenden als k -Nächste-Nachbarn-Imputation

bezeichnet. Auf Basis einer Faustregel wird dabei in der Voreinstellung stets $k = 0.01n$ verwendet; wenn nicht explizit anders angegeben ist dies auch der Wert, der in den Simulationsstudien in Kapitel 6 verwendet wird.

Der Vorteil bei einem solchen Vorgehen liegt vor allem darin, dass es im Gegensatz zu den anderen vorgestellten Imputationsmöglichkeiten verteilungsfrei ist und die Methodik somit auch für eine Vielzahl komplexer und untypischer Datensituationen programmiert und umgesetzt werden kann. Dies zeigt sich insbesondere auch in dem Anwendungsbeispiel aus Abschnitt 7.3, bei dem aufgrund der sehr geringen Stichprobengröße komplexe Regressionsimputationen, wie auch eine verteilungsbasierte Imputation auf Basis des EM-Algorithmus, keine Imputationen erzeugen können. Eben diese Verteilungsfreiheit führt aber auch dazu, dass keine (sinnvollen) multiplen Imputationen generiert werden können, was – wie zu Beginn dieses Kapitels beschrieben – meist zu einer Unterschätzung der Varianz führt. Für einige wenige, spezielle Situationen existieren jedoch Korrekturverfahren, die die Varianz nicht unterschätzen, vergleiche Chen und Shao (2001).

Problematisch erscheint ferner die Wahl einer geeigneten Metrik $d(\cdot, \cdot)$. Es ist sicher selten, dass ein Datensatz nur metrische Variablen enthält und die Wahl der euklidischen Distanz damit einen Sinn ergibt. Die Sensitivität der Ergebnisse bezüglich der Wahl der Metrik ist damit eine wichtige, jedoch in der Literatur in diesem Zusammenhang selten erwähnte, Problematik, die auch das oben angedeutete EMV-Paket von Gottardo (2008) nicht berücksichtigt. Es stellt als Alternative zur euklidischen Distanz nur die Pseudo-Metrik der Stichprobenkorrelation zur Verfügung.

Bootstrap-Imputation auf Basis des EM-Algorithmus

Die analytische Bestimmung oder eine iterative Approximation der prädiktiven a-posteriori-Verteilung (5.1) ist oft sehr aufwändig und zeitintensiv. Das *R*-Paket „Amelia II“ (Honaker, King und Blackwell (2008)) verfolgt einen neuen, wenig rechenintensiven, Bootstrap-basierten Ansatz um korrekte (multiple) Imputationen zu erhalten. Dabei wird der Datenmatrix D eine multivariate Normalverteilung unterstellt, $D \sim N(\mu, \Sigma)$. Kategoriale Größen mit k Ausprägungen werden dabei in $k - 1$ binäre Dummies umkodiert, binäre Größen bleiben bestehen. Wie zu Beginn dieses Kapitels erwähnt ist die vereinfachte Annahme einer multivariaten Normalverteilung nicht unproblematisch, liefert aber in vielen Situationen vernünftige Ergebnisse. Ausgehend von dieser Annahme werden M Bootstrap-Stichproben der gleichen Größe wie der Originaldaten gezogen und für jede Stichprobe auf Basis des EM-Algorithmus die posteriori Modi $\mu_{(m)}^*$ und $\Sigma_{(m)}^*$,

$m = 1, \dots, M$, bestimmt. Dieser Schritt soll die Betrachtung der posteriori-Verteilung $p(\theta|D^{\text{obs}})$ aus (5.1) simulieren und ersetzen. Eine ausführliche Rechtfertigung ohne formellen Beweis findet sich bei Honaker und King (2010). Um korrekte, multiple Imputationen $\tilde{x}_{ij}^{(m)}$, $m = 1, \dots, M$, gemäß der prädiktiven a-posteriori-Verteilung $p(D^{\text{mis}}|D^{\text{obs}})$ zu erhalten, können aufgrund der Annahme der multivariaten Normalverteilung die Werte schlicht über die daraus resultierende Regressionsvorhersage in den Originaldaten

$$\tilde{x}_{ij}^{(m)} = \mathbf{x}_{i/j} \beta_{(m)}^* + \epsilon_{(m)}^*$$

bestimmt werden. Dabei beschreibt $\mathbf{x}_{i/j}$ die i -te Beobachtung ohne den entsprechenden Wert der Variable X_j und $\beta_{(m)}^*$ bzw. $\epsilon_{(m)}^*$ sind die aus $\mu_{(m)}^*$ und $\Sigma_{(m)}^*$ eindeutig resultierenden Parameter. Somit erhält man M imputierte Datensätze, die als sinnvolle Wahl zur korrekten multiplen Imputation auf Basis von (5.1) angesehen werden können.

5.1.3 Weitere Ansätze

Es existieren in der Literatur drei weitere Ansätze zur Berücksichtigung der Problematik fehlender Daten in der Modellselektion, die allesamt Akaikes Informationskriterium für fehlende Daten adjustieren. Die erste Veröffentlichung von Shimodaira (1994), der sein Kriterium als PDIO (Predictive Divergence for Incomplete Observation models) bezeichnet, wie auch die darauf aufbauende Erweiterung von Cavanaugh und Shumway (1998), die ihre Modifikation mit AICcd (AIC complete data) benennen, betrachten ausschließlich den Fall fehlender Werte im Response und vollständig beobachteter Werte in den Kovariablen. Die vorliegende Arbeit diskutiert dagegen im Wesentlichen den Fall unvollständiger Werte in den Kovariablen, weswegen an dieser Stelle nur auf Cavanaugh und Shumway (1998, Abschnitt 4ff.) verwiesen werden soll; hier werden die Kernkonzepte der beiden Kriterien ausführlich motiviert, verglichen und ferner erläutert, wie die aufwändige Implementation von PDIO und AICcd gestaltet werden kann.

Einen etwas anderen Ansatz verfolgt die Arbeit von Claeskens und Consentino (2008). Ihr Kriterium ist eine Adjustierung des AIC und stellt ebenfalls einen unverzerrten Schätzer für die erwartete Kullback-Leibler-Distanz (3.21) dar. Voraussetzung für die Gültigkeit des Kriteriums ist dabei, dass die fehlenden Werte ausschließlich in den Kovariablen auftreten und nicht im Response. Es lautet

$$AIC_{\text{mis}} = -2Q(\hat{\theta}|\hat{\theta}) + 2K, \quad (5.15)$$

wobei $\hat{\theta}$ die ML-Schätzung von θ auf Basis des EM-Algorithmus beschreibt und $Q(\hat{\theta}|\hat{\theta})$ die dabei relevante Größe im E-Schritt darstellt, $Q(\hat{\theta}|\hat{\theta}) = \mathbb{E}_{D^{\text{mis}}|D^{\text{obs}}, \hat{\theta}} [\mathcal{L}(\hat{\theta}|D^{\text{mis}}, D^{\text{obs}})]$. Es ist offensichtlich, dass dies einen MAR-Fehlendmechanismus voraussetzt. Zur konkreten Berechnung von $Q(\hat{\theta}|\hat{\theta})$ verwenden die Autoren einen Monte-Carlo EM-Algorithmus, was insbesondere daran liegt, dass das AIC_{mis} nur dann Gültigkeit besitzt, wenn alle Modelle mindestens eine Kovariable mit fehlenden Werten enthalten; ansonsten muss eine Korrektur vorgenommen werden, die eine analytische Berechnung von $Q(\hat{\theta}|\hat{\theta})$ auch für vermeintlich einfache Verteilungsfamilien verhindert. Wie in Abschnitt 5.1.1 angedeutet, basieren die vorwiegend guten Ergebnisse dieses Kriteriums in den Simulationsstudien der Autoren auf der unhaltbaren Annahme, dass überangepasste Modelle als korrekt angesehen werden können.

Ein weiterer, offensichtlich trivialer Ansatz besteht darin, den Selektionsschritt auf dem Datensatz der vollständigen Fälle D_*^c durchzuführen. Für eine Menge an Kandidatenmodellen $\mathcal{M} = \{M_1 \dots, M_k\}$ wird dann dasjenige Modell $M_\kappa \in \mathcal{M}$ gewählt, das ein Kriterium oder Verfahren Γ , wie in Kapitel 3 beschrieben, auf D_*^c minimiert (bzw. maximiert) und somit zu der Wahl des Parameterschätzers

$$\hat{\theta}_{\text{CC}} = \hat{\theta}_\kappa : \arg \min_{M_\kappa \in \mathcal{M}} \{\Gamma(M_\kappa; \hat{\theta}_\kappa) | D_*^c\}, \quad \kappa = 1, \dots, k, \quad (5.16)$$

führt. Wie zu Beginn des Kapitels erläutert, hängt die Qualität dieses Schätzers sehr stark von dem zugrundeliegenden Fehlendmechanismus und der Reduzierung des Stichprobenumfangs ab. Die Simulationsergebnisse aus Kapitel 6 unterstreichen, dass Korrekturverfahren, wie etwa die Verwendung des AIC_W oder die „Selektion nach Imputation“-Strategie – wie zu erwarten – in der Regel bessere Schätzungen liefern als (5.16).

5.2 Modellmittelung bei fehlenden Daten

Die im vorangegangenen Abschnitt 5.1 angestellten Überlegungen erlauben es, die Problematik fehlender Daten in der Modellselektion sowohl unter Zuhilfenahme von Gewichtsansätzen (Abschnitt 5.1.1) als auch unter Verwendung geeigneter Imputationsmethoden (Abschnitt 5.1.2) zu berücksichtigen. Wie in Kapitel 4 ausführlich diskutiert führt die datengestützte Selektion von Modellen jedoch zu einer Unsicherheit, die in der Inferenz eigentlich nicht vernachlässigt werden sollte; dies gilt selbstverständlich auch im Kontext fehlender Daten. Im Folgenden wird daher aufgezeigt wie gängige frequentistische Modellmittelungsansätze für den Kontext fehlender Daten adjustiert werden

können. Die Eigenschaften der vorgeschlagenen Schätzer werden in Kapitel 6 ausführlich untersucht und bewertet.

5.2.1 Mittelung mit adjustierten Kriterien

Ein erster, pragmatischer Ansatz besteht darin, für fehlende Daten adjustierte Kriterien zur Konstruktion von Gewichten eines Modellmittelungsschätzers zu benutzen: Es liegt nahe, anstelle der exponentiellen SBC-Gewichte (4.4) bzw. der exponentiellen AIC-Gewichte (4.6), entsprechende exponentielle Gewichte auf Basis eines in diesem Kontext sinnhaften Kriteriums Γ zu verwenden, das explizit den Umstand fehlender Beobachtungen berücksichtigt. Wie bereits erwähnt, hält die Literatur für den Fall fehlender Responsewerte hierfür bisher das PDIO-Kriterium von Shimodaira (1994) und das AICcd von Cavanaugh und Shumway (1998), für fehlende Werte in den Kovariablen ein EM-basiertes Kriterium von Claeskens und Consentino (2008) sowie im Allgemeinen das gewichtete AIC-Kriterium AIC_W von Hens, Aerts und Molenberghs (2006) bereit. Für letzteres ergibt sich unter Beachtung der Konzepte aus den Abschnitten 4.2.1 und 5.1.1 der Modellmittelungsschätzer

$$\hat{\theta}_{AIC_W} = \sum_{\kappa=1}^k w_{\kappa}^{(8)} \hat{\theta}_{W,\kappa} \quad (5.17)$$

mit

$$w_{\kappa}^{(8)} = \frac{\exp(-\frac{1}{2}AIC_{W,\kappa})}{\sum_{\kappa=1}^k \exp(-\frac{1}{2}AIC_{W,\kappa})}, \quad (5.18)$$

wobei $\hat{\theta}_{W,\kappa}$ die gewichtete ML-Schätzung auf Basis der Gewichte (5.5) und $AIC_{W,\kappa}$ das gewichtete Akaike-Kriterium für das Modell $M_{\kappa} \in \mathcal{M}$ bezeichnet. Gemäß (4.12) kann die Varianz von $\hat{\theta}_{AIC_W}$ dabei über

$$\widehat{\text{Var}}(\hat{\theta}_{AIC_W}) = \left\{ \sum_{\kappa=1}^k w_{\kappa} \sqrt{\widehat{\text{Var}}(\hat{\theta}_{W,\kappa}|M_{\kappa}) + (\hat{\theta}_{W,\kappa} - \hat{\theta}_{AIC_W})^2} \right\}^2 \quad (5.19)$$

geschätzt werden. Damit wird die durch den Selektionsprozess verursachte Unsicherheit explizit berücksichtigt; wie bereits erwähnt erfordert die Schätzung der Gewichte für das AIC_W jedoch meist das Fitten eines GAMs, was die Wahl eines Glättungsparameters beinhaltet und zu einer weiteren Unsicherheit führt, die nicht durch (5.19) abgedeckt wird. Dies verdeutlichen auch die Simulationsergebnisse aus Kapitel 6.

5.2.2 Mittelung nach Imputation

Eine weitere Möglichkeit die Problematik fehlender Daten in der Modellmittelung zu berücksichtigen, besteht darin, den unvollständigen Datensatz D_* aufzufüllen und den daraus resultierenden imputierten Datensatz D^{imp} für die Bestimmung eines Modellmittelungsschätzers zu verwenden. Unabhängig von der gewählten Imputationsmethode ergibt sich für den Fall einer einfachen, nicht-multiplen Imputation der Schätzer

$$\hat{\theta}_{\text{imp}} = \sum_{\kappa=1}^k \{w_{\kappa} \hat{\theta}_{\kappa} | D^{\text{imp}}\}, \quad (5.20)$$

dessen Varianz gemäß (4.12) über

$$\widehat{\text{Var}}(\hat{\theta}_{\text{imp}}) = \left\{ \sum_{\kappa=1}^k w_{\kappa} \sqrt{\widehat{\text{Var}}(\hat{\theta}_{\kappa} | M_{\kappa}) + (\hat{\theta}_{\kappa} - \hat{\theta}_{\text{imp}})^2 | D^{\text{imp}}} \right\}^2 \quad (5.21)$$

geschätzt werden kann. In den Simulations- und Anwendungsbeispielen der Kapitel 6 und 7 werden hierfür die in Abschnitt 5.1.2 vorgestellten Imputationsprozeduren der kNN-Imputation, der verallgemeinerten Regressionsimputation (GLMRI, GAMRI) und der Bootstrap-basierten EM-Imputation verwendet. Für die Wahl der Gewichte in (5.20) ergeben sich dabei viele Möglichkeiten: Es können exponentielle Gewichte auf Basis des SBC, AIC oder jedes anderen sinnhaften Kriteriums Γ aus Kapitel 3, also

$$w_{\kappa}^{(9)} = \frac{\exp(-\frac{1}{2}\Gamma_{\kappa})}{\sum_{\kappa=1}^k \exp(-\frac{1}{2}\Gamma_{\kappa})}, \quad (5.22)$$

verwendet werden; auch Gewichte als Approximation an die a-posteriori-Wahrscheinlichkeit (4.3), die MMA-Gewichte (4.9) oder die OPT-Gewichte (4.11) ergeben einen Sinn. Aus diesen Überlegungen werden auch die großen Vorteile der Verwendung eines Schätzers auf Basis der imputierten Daten klar: Im Gegensatz zu (5.17) muss die Plausibilität eines Modells nicht über informationstheoretisch motivierte Kriterien beschrieben werden; nahezu jeder Ansatz aus Kapitel 3, etwa die Kriterien unter Beachtung der Vorhersagequalität oder auch der MDL-Methodik, eignen sich in diesem Zusammenhang um einen Schätzer zu konstruieren. In einem so weitreichenden Feld wie der Modellselektion und -mittelung erscheint dies von großem Vorteil.

Dennoch beachten weder (5.20) noch (5.21) die Unsicherheit bezüglich der Imputation. Es liegt insofern nahe, die Konzepte multipler Imputation, wie etwa in (5.2) und (5.3) beschrieben, zu adaptieren: Prinzipiell bedeutet dies nichts anderes als

1. M imputierte Datensätze durch Ziehen aus der prädiktiven a-posteriori-Verteilung (5.1) zu generieren,
2. auf jedem dieser M Datensätze den gewünschten Modellmittelungsschätzer zu berechnen und
3. die dadurch erhaltenen M Schätzer für eine finale Schätzung zu kombinieren.

Dadurch erhält man als eine sinnvolle Punktschätzung

$$\hat{\theta}_{\text{imp}}^M = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_{\text{imp}}^{(m)}, \quad (5.23)$$

wobei $\hat{\theta}_{\text{imp}}^{(m)}$ einen beliebigen Modellmittelungsschätzer auf Basis des m -ten imputierten Datensatzes beschreibt, $m = 1, \dots, M$. Die Varianz kann dabei offensichtlich über

$$\begin{aligned} \widehat{\text{Var}}(\hat{\theta}_{\text{imp}}^M) &= \frac{1}{M} \sum_{m=1}^M \widehat{\text{Var}}(\hat{\theta}_{\text{imp}}^{(m)}) + \frac{M+1}{M(M-1)} \sum_{m=1}^M (\hat{\theta}_{\text{imp}}^{(m)} - \hat{\theta}_{\text{imp}}^M)^2 \\ &= \frac{1}{M} \sum_{m=1}^M \left\{ \sum_{\kappa=1}^k w_{\kappa} \sqrt{\widehat{\text{Var}}(\hat{\theta}_{\kappa}^{(m)}) + (\hat{\theta}_{\kappa}^{(m)} - \hat{\theta}^{(m)})^2} \right\}^2 \\ &\quad + \frac{M+1}{M(M-1)} \sum_{m=1}^M (\hat{\theta}_{\text{imp}}^{(m)} - \hat{\theta}_{\text{imp}}^M)^2 \end{aligned} \quad (5.24)$$

geschätzt werden. Ein solches Vorgehen berücksichtigt sowohl die Unsicherheit bezüglich der Imputation als auch die Unsicherheit bezüglich des Selektionsschrittes und ist insofern als Weiterentwicklung sowohl von (4.12) als auch von (5.10) anzusehen.

Eine Reflexion über (5.23) und insbesondere (5.24) führt zu der Erkenntnis, dass das oben beschriebene Prozedere eine Rangordnung in der Betrachtung der Unsicherheitskomponenten impliziert: Zuerst wird die Unsicherheit bezüglich der Modellwahl über Bestimmung eines FMA-Schätzers berücksichtigt, anschließend die Unsicherheit bezüglich der Imputation über die Verwendung multipler Imputationen. Auch wenn ein solches Vorgehen gemäß den oben beschriebenen drei Schritten statistischer Praxis entsprechen mag und sicherlich auch plausibel ist, so stellt sich dennoch die Frage, wie sich eine Änderung dieser Reihenfolge auf den Schätzer (5.23) und seine Varianz auswirkt und ob damit sinnvolle Schätzungen generiert werden können oder nicht. Berechnet man also

zuerst einen Schätzer, der die Imputationsunsicherheit berücksichtigt und anschließend die Unsicherheit bezüglich der Modellwahl, so erhält man offensichtlich

$$\hat{\theta}_{\text{imp,alt}}^M = \sum_{\kappa=1}^k w_{\kappa}^M \hat{\theta}_{\kappa}^M, \quad (5.25)$$

wobei

$$w_{\kappa}^M = \frac{1}{M} \sum_{m=1}^M w_{\kappa}^{(m)}. \quad (5.26)$$

Tatsächlich sind die Punktschätzungen $\hat{\theta}_{\text{imp,alt}}^M$ und $\hat{\theta}_{\text{imp}}^M$ in der Regel nicht identisch, da

$$\begin{aligned} \hat{\theta}_{\text{imp,alt}}^M &= \sum_{\kappa=1}^k w_{\kappa}^M \hat{\theta}_{\kappa}^M = \frac{1}{M^2} \sum_{\kappa=1}^k \left\{ \sum_{m=1}^M w_{\kappa}^{(m)} \sum_{m=1}^M \hat{\theta}_{\kappa}^{(m)} \right\} \\ &\stackrel{\text{i.A.}}{\neq} \frac{1}{M} \sum_{m=1}^M \sum_{\kappa=1}^k w_{\kappa}^{(m)} \hat{\theta}_{\kappa}^{(m)} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_{\text{imp}}^{(m)} = \hat{\theta}_{\text{imp}}^M. \end{aligned}$$

Auch die $\hat{\theta}_{\text{imp,alt}}^M$ zugehörige Varianzschätzung, die über

$$\widehat{\text{Var}}(\hat{\theta}_{\text{imp,alt}}^M) = \left\{ \sum_{\kappa=1}^k w_{\kappa}^M \sqrt{\widehat{\text{Var}}(\hat{\theta}_{\kappa,\text{imp}}^M) + (\hat{\theta}_{\text{imp}}^M - \hat{\theta}_{\kappa,\text{imp}}^M)^2} \right\}^2 \quad (5.27)$$

erhalten werden kann, wobei

$$\widehat{\text{Var}}(\hat{\theta}_{\kappa,\text{imp}}^M) = \frac{1}{M} \sum_{m=1}^M \widehat{\text{Var}}(\hat{\theta}_{\kappa,\text{imp}}^{(m)}) + \frac{M+1}{M(M-1)} \sum_{m=1}^M (\hat{\theta}_{\kappa,\text{imp}}^M - \hat{\theta}_{\kappa,\text{imp}}^{(m)})^2,$$

stimmt nicht mit der Varianzschätzung für $\hat{\theta}_{\text{imp}}^M$ gemäß (5.24) überein. Der Vergleich von (5.24) und (5.27) ist dabei keineswegs trivial, da die Quadrate größerer, unterschiedlich konzipierter Summen verglichen werden müssen. Entscheidend ist der Unterschied in Bezug auf die Einflussnahme der Gewichte $w_{\kappa}^{(m)}$: Während bei den „klassischen“ Schätzungen nach (5.23) und (5.24) die Gewichte des m -ten Datensatzes auch nur auf die Schätzungen des entsprechenden m -ten Datensatzes wirken können, so ist dies bei den „alternativen“ Schätzungen nach (5.25) und (5.27) nicht der Fall; Gewichte, die für den m -ten Datensatz berechnet werden, wirken indirekt über (5.26) immer auch

auf die Schätzungen der anderen, imputierten Datensätze. Ein einfaches Beispiel soll dies illustrieren. Betrachtet man $k = 2$ Modelle und $M = 2$ Imputationen, so erhält man als entsprechende Schätzungen für θ :

$$\begin{aligned}\hat{\theta}_{\text{imp}}^M &= \frac{1}{M} \sum_{m=1}^M \hat{\theta}_{\text{imp}}^{(m)} = \frac{1}{M} \sum_{m=1}^M \sum_{\kappa=1}^k w_{\kappa}^{(m)} \hat{\theta}_{\kappa}^{(m)} \\ &= \frac{1}{2} w_1^{(1)} \hat{\theta}_1^{(1)} + \frac{1}{2} w_2^{(1)} \hat{\theta}_2^{(1)} + \frac{1}{2} w_1^{(2)} \hat{\theta}_1^{(2)} + \frac{1}{2} w_2^{(2)} \hat{\theta}_2^{(2)}\end{aligned}$$

bzw.

$$\begin{aligned}\hat{\theta}_{\text{imp,alt}}^M &= \sum_{\kappa=1}^k w_{\kappa}^M \hat{\theta}_{\kappa}^M = \frac{1}{M^2} \sum_{\kappa=1}^k \left\{ \sum_{m=1}^M w_{\kappa}^{(m)} \sum_{m=1}^M \hat{\theta}_{\kappa}^{(m)} \right\} \\ &= \frac{1}{4} (w_1^{(1)} + w_1^{(2)}) (\hat{\theta}_1^{(1)} + \hat{\theta}_1^{(2)}) + \frac{1}{4} (w_2^{(1)} + w_2^{(2)}) (\hat{\theta}_2^{(1)} + \hat{\theta}_2^{(2)}) \\ &= \frac{1}{4} w_1^{(1)} \hat{\theta}_1^{(1)} + \frac{1}{4} w_1^{(1)} \hat{\theta}_1^{(2)} + \frac{1}{4} w_1^{(2)} \hat{\theta}_1^{(1)} + \frac{1}{4} w_1^{(2)} \hat{\theta}_1^{(2)} \\ &\quad + \frac{1}{4} w_2^{(1)} \hat{\theta}_2^{(1)} + \frac{1}{4} w_2^{(1)} \hat{\theta}_2^{(2)} + \frac{1}{4} w_2^{(2)} \hat{\theta}_2^{(1)} + \frac{1}{4} w_2^{(2)} \hat{\theta}_2^{(2)}.\end{aligned}$$

Offensichtlich wirken bei $\hat{\theta}_{\text{imp,alt}}^M$ Gewichte des m -ten Datensatzes auch auf Schätzungen des nicht- m -ten Datensatzes, so beispielsweise $w_1^{(1)}$, das auf $\hat{\theta}_1^{(2)}$ wirkt.

Die Simulationsergebnisse aus Abschnitt 6.3 werden bestätigen, dass – wie vermutet – die Verwendung der vorgestellten, alternativen Schätzungen in den betrachteten Situationen tatsächlich zu qualitativ schlechteren Ergebnissen führt als die intuitiven Schätzungen nach (5.23) und (5.24). Die Reihenfolge der Betrachtung der Unsicherheitskomponenten ist bei Gebrauch multipler Imputationen im Kontext der Modellmittelung also entscheidend.

6. Simulationsstudien

Um die Eigenschaften und das Verhalten der vorgeschlagenen Punktschätzungen und deren Varianz zu eruieren, werden in diesem Kapitel ausführliche Monte-Carlo-Simulationen durchgeführt. Dabei sind die folgenden Fragestellungen von besonderem Interesse:

1. Sind Modellmittlungsverfahren der Modellselektion generell überlegen?
2. Ist ein pragmatischer Modellmittlungsansatz unter Verwendung kriteriumsbasierter Gewichte einem Ansatz der unter Optimalitätsbetrachtungen konstruiert wurde, so etwa dem MMA-Schätzer von Hansen, zwingenderweise unterlegen?
3. Welche Erfolgsaussichten versprechen die in Kapitel 5 vorgestellten Korrekturverfahren für fehlende Daten?

In den Abschnitten 6.1 und 6.2 werden hierfür insbesondere die Qualität der entsprechenden Punktschätzungen untersucht und bewertet; Abschnitt 6.3 widmet sich hauptsächlich der Qualität der zugehörigen Varianzschätzungen und den Auswirkungen multipler Imputation. Abschnitt 6.4 fasst die wichtigsten Resultate zusammen.

6.1 Lineare Regression

In diesem Abschnitt werden am Beispiel zweier ausgewählter FMA-Schätzer sowie eines FMS-Schätzers die oben angeführten Fragestellungen im Kontext des linearen Regressionsmodells

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I),$$

unter Beachtung der Problematik fehlender Daten erörtert, $\beta = (\alpha, \gamma)'$. Betrachtet werden $\mathcal{R} = 500$ Simulationsläufe, wobei für jeden Lauf je ein $n_{\text{tr}} \times p$ Trainingsdatensatz und ein $n_{\text{test}} \times p$ Testdatensatz erzeugt wird mit $n_{\text{tr}} = 450$, $n_{\text{test}} = 50$ und $p = 6$.

Mit Hilfe eines Clayton-Copulas wird eine multivariate Verteilung für die insgesamt fünf Kovariablen mit den Randverteilungen $X_1 \sim N(0.5, 1)$, $X_2 \sim \log N(0.5, 0.5)$, $X_3 \sim Weibull(1.75, 1.9)$, $X_4 \sim B(1, 0.3)$ und $X_5 \sim Ga(0.25, 2)$ generiert. Die Verteilungen sind so gewählt, dass die Varianz der Kovariablen aus Gründen der Standardisierung immer in etwa Eins beträgt; der Copula-Parameter beträgt durchgehend $\theta_{\text{cop}} = 1$ und resultiert damit in einer mittelstarken Korrelation zwischen allen Kovariablen; für Details bezüglich der Verwendung von Copulas in der statistischen Software *R* sei auf Yan (2007) verwiesen. Der Response y wird aus einer Normalverteilung $N(\mu_0, \sigma_0)$ gezogen mit $\mu_0 = 2.5 - 3X_1 - 0.3X_2 - 2X_4$ und $\sigma_0 = \exp(1)$. Dies bedeutet, dass der Erwartungswert von y maßgeblich von den Variablen X_1 , X_2 und X_4 bestimmt wird, daher gilt $\alpha_{\text{wahr}} = 2.5$ und $\gamma_{\text{wahr}} = (-3, -0.3, 0, -2, 0)'$ und damit $\beta_{\text{wahr}} = (\alpha_{\text{wahr}}, \gamma'_{\text{wahr}})'$; der komplette Datensatz wird als $D_{\text{sim1}} = \{y, X_1, X_2, X_3, X_4, X_5\}$ bezeichnet und wird in jedem Simulationslauf neu generiert. Anschließend werden unter Verwendung eines MAR-Fehlendmechanismus Werte von X_1 , X_4 und X_5 gemäß der Fehlwahrscheinlichkeitsfunktionen

$$\begin{aligned} \pi_{X_1}(y) &= 1 - \frac{1}{0.04y^2 + 1}, & \pi_{X_4}(X_2) &= 1 - \frac{1}{1 + 0.02X_2^3}, \\ \pi_{X_5}(X_3) &= 1 - \frac{1}{1 + \exp\{1 - 2X_3\}}, \end{aligned}$$

als fehlend deklariert. Bei den $\mathcal{R} = 500$ Simulationsläufen fehlen damit im Mittel 31% der Werte von X_1 , 15% der Werte von X_4 und 18% der Werte von X_5 . Die Gestalt der Fehlwahrscheinlichkeitsfunktionen ist in Abbildung 6.1 illustriert. Um den Zusammenhang

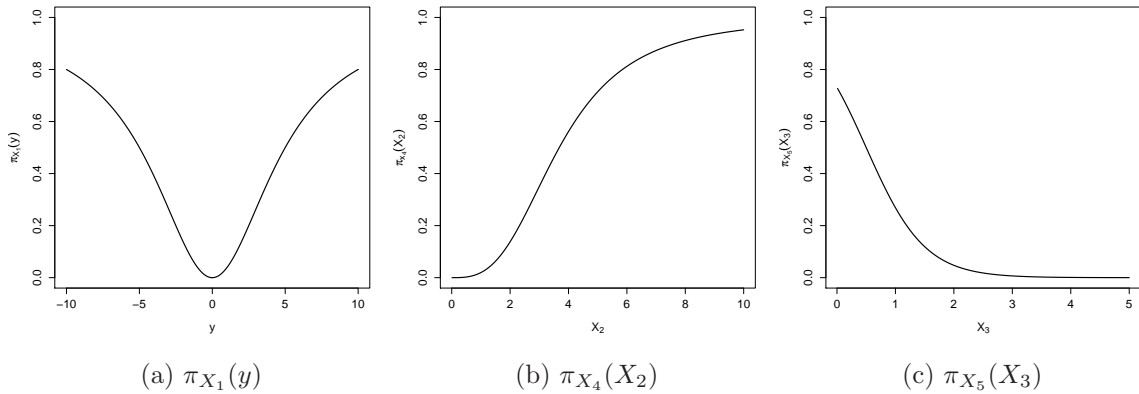


Abb. 6.1: Gestalt der verwendeten Fehlwahrscheinlichkeitsfunktionen

von y und den Kovariablen zu modellieren, werden fünf konkurrierende Kandidatenmodelle¹⁹ betrachtet:

$$M_1 : y = \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 ,$$

$$M_2 : y = \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 ,$$

$$M_3 : y = \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_4 X_4 + \gamma_5 X_5 ,$$

$$M_4 : y = \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_4 X_4 ,$$

$$M_5 : y = \alpha + \gamma_1 X_1 + \gamma_4 X_4 .$$

Offensichtlich ist das Modell M_4 , das die Kovariablen X_1 , X_4 und X_5 enthält, am besten geeignet, um den oben formulierten, datengenerierenden Prozess zu beschreiben. Um nun sowohl die Effekte der Modellselektionsunsicherheit als auch die Problematik fehlender Daten zu berücksichtigen, werden 20 Schätzer, wie in Tabelle 6.1 aufgelistet, analysiert. Alle FMA-Schätzer besitzen dabei die Form

$$\hat{\beta}^* = \sum_{\kappa=1}^k w_{\kappa}^* \hat{\beta}_{\kappa}^* \quad (6.1)$$

und unterscheiden sich sowohl bezüglich der Datengrundlage zur Berechnung von $\hat{\beta}_{\kappa}^*$ als auch bezüglich der Wahl der Gewichte w_{κ}^* . Der verwendete FMS-Schätzer kann dabei als Spezialfall des FMA-Schätzers (6.1) angesehen werden, da der durch das Kriterium $\Gamma = \text{AIC}$ ausgewählte Schätzer $\hat{\beta}_{\kappa}^{\text{AIC}}$ das Gewicht 1 zugewiesen bekommt, alle weiteren Schätzer das Gewicht 0.

Die Qualität dieser Schätzer wird anhand vier verschiedener Verlustfunktionen beurteilt: Die Verlustfunktion L_1 beschreibt dabei den MSE der angeführten Schätzer bezüglich β_{wahr} ,

$$L_1 = \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} \left\{ \sum_{j=1}^p (\hat{\beta}_{j,r}^* - \beta_{j,\text{wahr}})^2 \right\} , \quad (6.2)$$

¹⁹ Die Auswahl der Kandidatenmodelle bildet die Grundlage zur Berechnung der FMA- und FMS-Schätzer gemäß Tabelle 6.1. Einzige Ausnahme ist der Mallows-Model-Averaging-Schätzer, der gemäß seiner Definition (vgl. Abschnitt 4.2.2) als Kandidatenmodelle *immer* das volle Modell (hier: M_1) und die entsprechenden verschachtelten Submodelle auf Basis der geordneten Regressoren verwendet. Dies gilt es auch für die folgenden Experimente in diesem Abschnitt zu beachten, da an den entsprechenden Stellen auf diesen Sachverhalt nicht noch einmal explizit darauf hingewiesen wird.

Auswahl an FMA- und FMS-Schätzern $\hat{\beta}^*$ bei fehlenden Daten

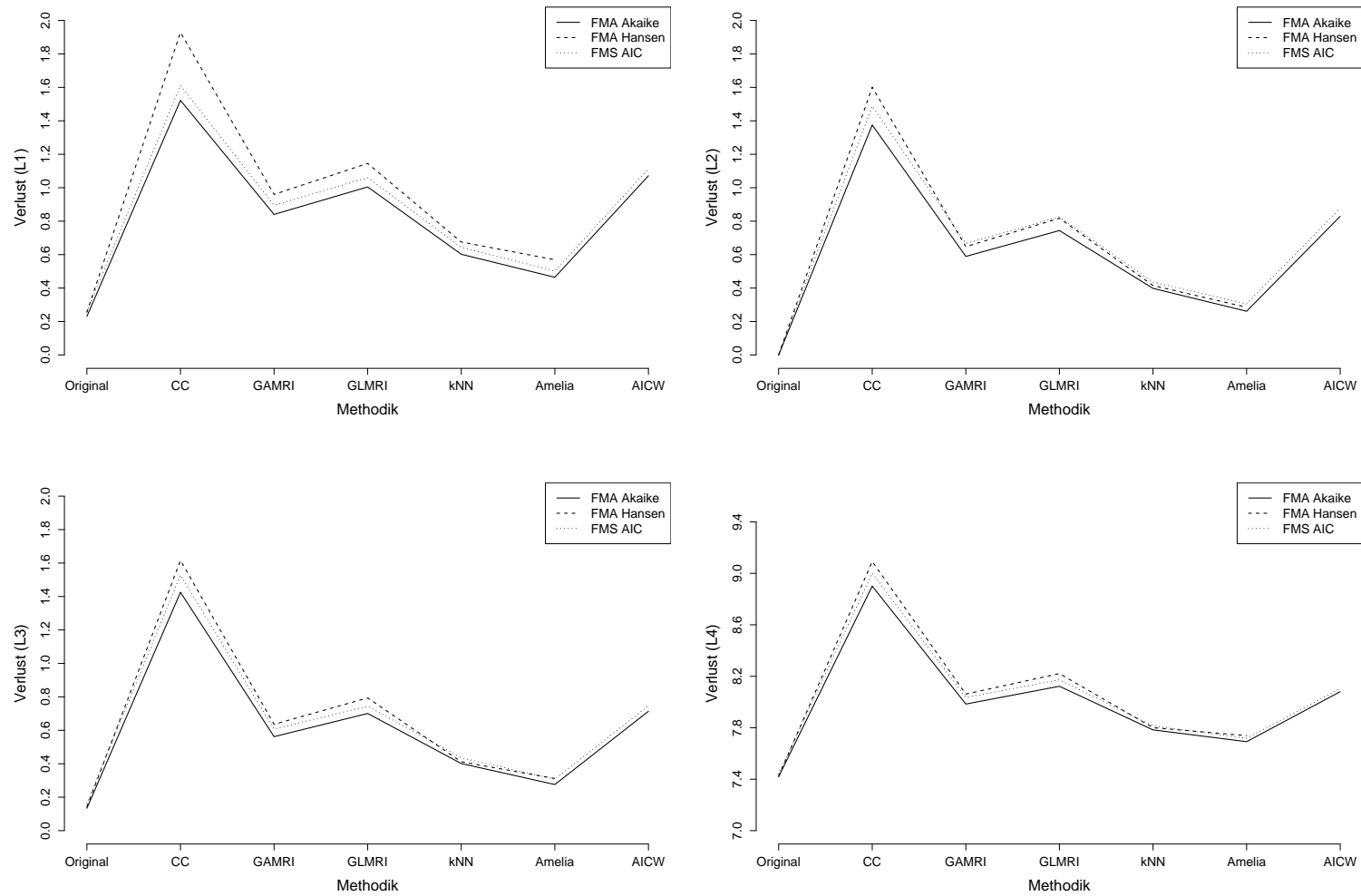
- (a) **FMA-Akaike-Schätzer:** der FMA-Schätzer auf Basis der exponentiellen AIC-Gewichte (4.6) bzw. (5.18), der den Umstand fehlender Daten wie folgt berücksichtigt:
- 1) *Original* – der FMA-Akaike-Schätzer unter Verwendung des vollständigen Datensatzes D_{sim1} ; dieser Schätzer dient als Referenz.
 - 2) *CC* – der FMA-Akaike-Schätzer unter Verwendung des Subdatensatzes der vollständig beobachteten Fälle D_* .
 - 3) *GAMRI*, *GLMRI*, *kNN*, *Amelia* – es wird das Prinzip „Mittelung nach Imputation“ angewendet; dies entspricht dem Schätzer (5.20) unter Verwendung des aufgefüllten Datensatzes D^{imp} gemäß den Imputationsmethoden der in Abschnitt 5.1.2 vorgestellten verallgemeinerten Regressionsimputation (*GAMRI*, *GLMRI*), der k-Nächsten-Nachbarn-Imputation (*kNN*) und der Bootstrap-basierten Imputation des R-Pakets „Amelia II“ (*Amelia*).
 - 4) AIC_W – entspricht dem adjustierten FMA-Schätzer (5.17).
- (b) **FMA-Hansen-Schätzer:** der FMA-Schätzer auf Basis der optimalen Gewichte (4.9), der den Umstand fehlender Daten wie in (a) gemäß der Strategien 1), 2) und 3) berücksichtigt; dieser Schätzer wird auch als MMA-Schätzer bezeichnet.
- (c) **FMS-AIC-Schätzer:** der FMS-Schätzer für den $\Gamma = \text{AIC}$ und der den Umstand fehlender Daten wie folgt berücksichtigt:
- 1) *Original* – der FMS-AIC-Schätzer wird für den vollständigen Datensatz D_{sim1} berechnet.
 - 2) *CC* – entspricht dem FMS-Schätzer (5.16).
 - 3) *GAMRI*, *GLMRI*, *kNN*, *Amelia* – es wird das Prinzip „Selektion nach Imputation“ angewendet; der Schätzer (5.8) basiert dabei auf den Imputationsmethoden wie in (a) beschrieben.
 - 4) AIC_W – der Schätzer wird durch das Kriterium $\Gamma = \text{AIC}_W$ bestimmt.
-

Tab. 6.1: Die im Grundszenario verwendeten Modellmittelungsschätzer und Modellselektionsschätzer für verschiedene Strategien zur Berücksichtigung der Problematik fehlender Daten

wobei $\hat{\beta}_{j,r}^*$ eine FMA- bzw. FMS-Schätzung des j -ten Elements von β , $j = 1 \dots, p$, im r -ten Simulationslauf bezeichnet; ferner wird der MSE bezüglich der r -ten Schätzung auf Basis der Originaldaten, $\hat{\beta}_r^{\text{org}}$, betrachtet,

$$L_2 = \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} \left\{ \sum_{j=1}^p (\hat{\beta}_{j,r}^* - \hat{\beta}_{j,r}^{\text{org}})^2 \right\}, \quad (6.3)$$

wobei sich $\hat{\beta}_r^{\text{org}}$ immer auf die entsprechende FMA- bzw. FMS-Methodik bezieht. Um die Vorhersagequalität zu beurteilen, wird der MSE der Schätzer bezüglich μ im Testdatensatz berechnet,

Abb. 6.2: Resultate im Grundszenario bezüglich der Verlustfunktionen L_1 , L_2 , L_3 und L_4

stets $H_1 : F_X < F_Y$. Es zeigt sich, dass – auch wenn die Unterschiede gering sind – die Verluste \tilde{L}_3 für den FMA-Akaike-Schätzer signifikant unter denen des FMS-AIC-Schätzers liegen bzw. für die GLMRI-Methodik signifikant geringer sind als für die Complete Case Analyse. Für die Verluste \tilde{L}_8 gilt dasselbe mit Ausnahme der FMS-AIC- und FMA-Akaike-Schätzungen für die GAMRI-Methodik, hier wird die Nullhypothese beibehalten. Die oben getroffenen Grundaussagen können also in weiten Teilen über alle Simulationsläufe hinweg bestätigt werden.

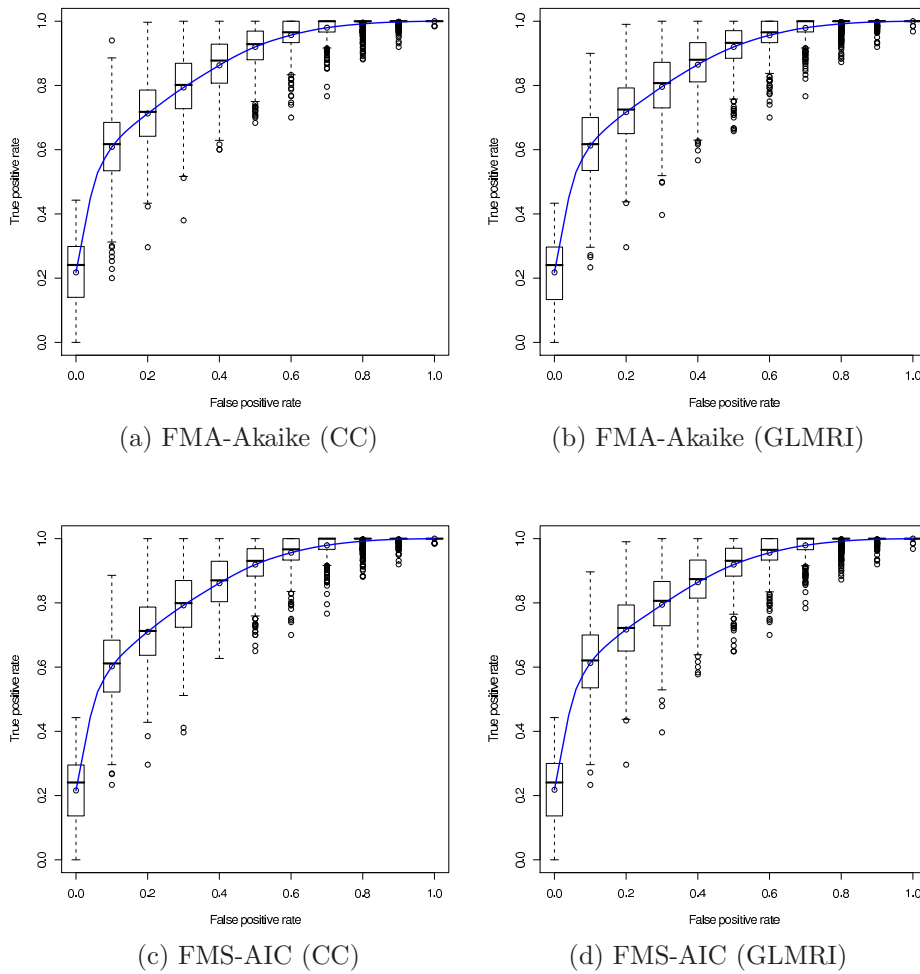


Abb. 6.6: Veranschaulichung des Verlustes \tilde{L}_8 für einige ausgewählte Schätzer über die entsprechenden ROC-Kurven

Um nun die Qualität der Varianzschätzungen der FMA- bzw. FMS-Schätzer zu beurteilen, wird erneut der empirische Standardfehler (6.7) für jedes β_j , $j = 0, \dots, 5$, berechnet

	se(β_0)		se(β_1)		se(β_2)		se(β_3)		se(β_4)		se(β_5)		se(β_6)	
FMA-Akaike-Schätzer														
1) Original	0.29	(0.30)	0.16	(0.16)	0.14	(0.15)	0.07	(0.08)	0.39	(0.41)	0.07	(0.08)	0.31	(0.33)
2) CC	0.39	(0.39)	0.23	(0.24)	0.19	(0.20)	0.07	(0.09)	0.49	(0.51)	0.07	(0.09)	0.42	(0.46)
3) GAMRI	0.28	(0.34)	0.17	(0.22)	0.13	(0.17)	0.09	(0.12)	0.41	(0.55)	0.08	(0.15)	0.33	(0.45)
GLMRI	0.28	(0.32)	0.17	(0.21)	0.13	(0.16)	0.08	(0.12)	0.41	(0.53)	0.08	(0.14)	0.33	(0.40)
kNN	0.30	(0.35)	0.18	(0.21)	0.14	(0.17)	0.08	(0.10)	0.43	(0.50)	0.08	(0.11)	0.36	(0.45)
Amelia	0.30	(0.36)	0.17	(0.19)	0.14	(0.18)	0.08	(0.10)	0.40	(0.47)	0.08	(0.13)	0.33	(0.35)
MI	0.34	(0.34)	0.19	(0.18)	0.16	(0.16)	0.09	(0.10)	0.46	(0.43)	0.10	(0.11)	0.38	(0.32)
4) AIC _W	0.41	(0.57)	0.23	(0.29)	0.20	(0.32)	0.12	(0.20)	0.53	(0.63)	0.11	(0.19)	0.44	(0.54)
FMA-Hansen-Schätzer														
1) Original	0.35	(0.30)	0.17	(0.17)	0.15	(0.14)	0.14	(0.13)	0.61	(0.44)	0.12	(0.11)	0.41	(0.35)
2) CC	0.49	(0.41)	0.24	(0.23)	0.22	(0.20)	0.16	(0.14)	0.67	(0.55)	0.11	(0.12)	0.40	(0.42)
3) GAMRI	0.33	(0.35)	0.18	(0.22)	0.15	(0.16)	0.14	(0.15)	0.60	(0.59)	0.12	(0.16)	0.37	(0.44)
GLMRI	0.32	(0.35)	0.18	(0.21)	0.15	(0.16)	0.14	(0.15)	0.61	(0.57)	0.12	(0.16)	0.37	(0.39)
kNN	0.37	(0.36)	0.19	(0.21)	0.16	(0.15)	0.14	(0.14)	0.65	(0.54)	0.12	(0.14)	0.43	(0.45)
Amelia	0.37	(0.37)	0.18	(0.20)	0.16	(0.17)	0.14	(0.14)	0.60	(0.50)	0.12	(0.15)	0.38	(0.35)
MI	0.40	(0.35)	0.20	(0.19)	0.17	(0.16)	0.15	(0.14)	0.65	(0.46)	0.14	(0.13)	0.43	(0.31)
FMS-AIC-Schätzer														
1) Original	0.25	(0.32)	0.15	(0.17)	0.11	(0.16)	0.02	(0.11)	0.37	(0.41)	0.02	(0.10)	0.30	(0.33)
2) CC	0.30	(0.44)	0.22	(0.25)	0.10	(0.24)	0.01	(0.11)	0.45	(0.54)	0.02	(0.12)	0.32	(0.51)
3) GAMRI	0.24	(0.36)	0.16	(0.23)	0.09	(0.18)	0.04	(0.15)	0.39	(0.55)	0.03	(0.17)	0.29	(0.46)
GLMRI	0.23	(0.35)	0.16	(0.21)	0.08	(0.17)	0.03	(0.15)	0.39	(0.53)	0.04	(0.17)	0.29	(0.40)
kNN	0.26	(0.39)	0.17	(0.22)	0.11	(0.18)	0.03	(0.13)	0.41	(0.51)	0.03	(0.14)	0.32	(0.45)
Amelia	0.26	(0.40)	0.16	(0.20)	0.11	(0.19)	0.03	(0.13)	0.38	(0.47)	0.03	(0.16)	0.29	(0.36)
MI	0.32	(0.36)	0.19	(0.18)	0.15	(0.17)	0.06	(0.12)	0.45	(0.43)	0.08	(0.13)	0.36	(0.32)
4) AIC _W	0.38	(0.59)	0.22	(0.29)	0.17	(0.32)	0.07	(0.22)	0.52	(0.64)	0.06	(0.21)	0.41	(0.55)

Tab. 6.8: Mittlere geschätzte Standardfehler unter Verwendung von (6.16); in Klammern die zugehörigen empirischen Standardfehler nach (6.7)

Alternative Varianzschätzungen

An dieser Stelle soll noch einmal die Diskussion aus Abschnitt 5.2.2 aufgenommen werden. Es stellt sich die Frage, ob die Reihenfolge der Berücksichtigung der Unsicherheiten eine Rolle spielt oder nicht. In den vorangegangenen beiden Experimenten wurde gemäß gängiger statistischer Konvention zuerst jede Form von Inferenz auf den M imputierten Datensätzen durchgeführt, anschließend wurden diese Schätzungen kombiniert. Dies impliziert, dass zuerst die Modellselektionsunsicherheit berücksichtigt wird, danach die Imputationsunsicherheit. Die Verwendung der Ideen von (5.27) führt dagegen für die vorliegenden beiden Experimente zu folgender, alternativen Schätzung des mittleren Standardfehlers:

$$\widehat{\text{se}}_{\beta_j} = \frac{1}{\mathcal{R}} \sum_{r=1}^{\mathcal{R}} \left\{ \sum_{\kappa=1}^k w_{\kappa}^M \sqrt{\widehat{\text{Var}}(\hat{\beta}_{j,\kappa,r}^{*,M}) + (\hat{\beta}_{j,r}^{*,M} - \hat{\beta}_{j,\kappa,r}^{*,M})^2} \right\}. \quad (6.17)$$

Dabei ist

$$\widehat{\text{Var}}(\hat{\beta}_{j,\kappa,r}^{*,M}) = \frac{1}{M} \sum_{m=1}^M \widehat{\text{Var}}(\hat{\beta}_{j,\kappa,r}^{*,(m)}) + \frac{M+1}{M(M-1)} \sum_{m=1}^M (\hat{\beta}_{j,\kappa,r}^{*,M} - \hat{\beta}_{j,\kappa,r}^{*,(m)})^2, \quad (6.18)$$

wobei $\hat{\beta}_{j,\kappa,r}^{*,M} = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_{j,\kappa,r}^{*,(m)}$ und $w_{\kappa}^M = \frac{1}{M} \sum_{m=1}^M w_{\kappa}^{(m)}$. Offensichtlich wird dadurch zuerst für jedes Modell ein Schätzer für den Standardfehler berechnet, der die Unsicherheit bezüglich der Imputationen berücksichtigt und dann anschließend über die Modelle gewichtet mittelt.

Im Folgenden werden nun erneut die beiden vorangegangenen Experimente durchgeführt und die mittleren geschätzten Standardfehler nach (6.16) den alternativen Schätzungen gemäß (6.17) gegenübergestellt. Um die Schätzungen weiter zu stabilisieren, werden die Simulationsläufe erhöht: $\mathcal{R} = 1000$. Betrachtet wird ausschließlich der FMA-Akaike-Schätzer. Die wesentlichen Resultate dafür sind in den Tabellen 6.10 und 6.11 zusammengefasst.

Es ist zu erkennen, dass für das erste Experiment die beiden Schätzungen nicht allzu weit voneinander entfernt liegen. Im Trend scheint die alternative Schätzung nach (6.17) die Standardfehler jedoch etwas stärker zu unter- bzw. überschätzen. Für die Resultate bezüglich des zweiten Experiments ist dies noch deutlicher zu erkennen: Während die geschätzten Standardfehler für die Effekte X_1 und X_2 noch relativ ähnlich eingeschätzt

	se(β_0)		se(β_1)		se(β_2)		se(β_3)	
MI	0.34	(0.34)	0.19	(0.19)	0.16	(0.17)	0.09	(0.10)
MI alt.	0.36	(0.34)	0.20	(0.19)	0.15	(0.17)	0.07	(0.10)

	se(β_4)		se(β_5)		se(β_6)	
MI	0.47	(0.41)	0.10	(0.11)	0.38	(0.32)
MI alt.	0.49	(0.41)	0.08	(0.11)	0.39	(0.32)

Tab. 6.10: Ergebnisse für die mittlere klassische Varianzschätzung (6.16) und die mittlere alternative Varianzschätzung (6.17) des FMA-Akaike-Schätzers im Grundszenario; in Klammern die zugehörigen empirischen Standardfehler nach (6.7)

	se(β_0)		se(β_1)		se(β_2)		se(β_3)	
MI	0.34	(0.34)	0.17	(0.17)	0.18	(0.19)	0.08	(0.09)
MI alt.	0.37	(0.34)	0.18	(0.17)	0.18	(0.19)	0.06	(0.09)

	se(β_4)		se(β_5)		se(β_6)	
MI	0.41	(0.42)	0.12	(0.13)	0.33	(0.35)
MI alt.	0.46	(0.42)	0.09	(0.12)	0.42	(0.35)

Tab. 6.11: Ergebnisse für die mittlere klassische Varianzschätzung (6.16) und die mittlere alternative Varianzschätzung (6.17) des FMA-Akaike-Schätzers in Experiment 2; in Klammern die zugehörigen empirischen Standardfehler nach (6.7)

werden, so werden sie für X_3 und X_5 von (6.17) im Vergleich zu (6.16) merklich unterschätzt. Noch größer gestaltet sich Unterschied für die bereits als kritisch eingeordneten Effekte X_4 und X_1X_4 ; die alternative Schätzung überschätzt die Standardfehler hier deutlich.

Während (6.16) für das zweite Experiment eine adäquate Modellierung der Varianz bietet, ist dies bei (6.17) nicht zu erkennen. Insgesamt scheint das klassische Vorgehen, also zuerst die Modellselektionsunsicherheit zu beachten, dann Imputationsunsicherheit, wie erwartet die bessere Wahl zu sein.

6.4 Zusammenfassung

Die in diesem Kapitel gewonnenen Erkenntnisse sind äußerst vielfältig und in vielerlei Hinsicht aufschlussreich für die gewählten Modellselektions- und Modellmittelungsverfahren im Kontext fehlender Daten. Auch wenn Simulationsergebnisse grundsätzlich immer nur mit Vorsicht auf Situationen außerhalb der betrachteten übertragen werden können und die Resultate zu einem gewissen Teil von den gewählten Einstellungen abhängen, so ergeben sich für die aufwändig und intensiv durchgeführten Studien dieses Kapitels dennoch einige entscheidende Feststellungen:

- Unabhängig von der gewählten Methodik zur Berücksichtigung fehlender Werte sind für den FMA-Akaike-Schätzer fast durchgängig bessere Punktschätzungen als beim FMS-AIC- bzw. FMA-Hansen-Schätzer zu erkennen. Letzterer wirkt insgesamt sehr instabil; die Qualität seiner Schätzungen hängt stark von der betrachteten Situation ab. In einigen Fällen, etwa bei den Experimenten 2 und 3 in Abschnitt 6.1, liefert er zwar insbesondere für die Verlustfunktionen L_2 , L_3 und L_4 bessere Schätzungen als der FMS-AIC-Schätzer, insgesamt kann jedoch nicht konstatiert werden, dass er dem Selektionsschätzer überlegen ist. Die Qualität der entsprechenden Varianzschätzungen scheint beim FMA-Akaike-Schätzer ebenfalls am besten zu sein.
- Die Verwendung der vollständigen Fälle (CC) führt bei allen FMA- und FMS-Schätzern nahezu ausschließlich zu sehr unzureichenden Punktschätzungen – sowohl bei Betrachtung der linearen Regression als auch bei Betrachtung der logistischen Regression. Dennoch sind die zugehörigen Varianzschätzungen in einem akzeptablen Bereich; sie unterschätzen die empirische Varianz nur in etwas komplexeren Situationen und auch dann nur marginal.
- Die Verwendung des AIC_W , entweder direkt zur Selektion des Modells oder alternativ zur Konstruktion von Modellmittelungsgewichten, führt in Abschnitt 6.1 (mit Ausnahme von Experiment 7) in der Regel zu besseren Schätzungen als eine Complete Case Analyse und zu etwas schlechteren Ergebnissen als die Imputationsverfahren. In den betrachteten Situationen der logistischen Regressionsanalyse aus Abschnitt 6.2, kann die AIC_W -Methodologie meist keinen Gewinn erbringen. Sie schneidet häufig am schlechtesten ab. Die Varianz der entsprechenden Schätzer wird in allen Simulationen der drei vorangegangenen Abschnitte 6.1-6.3 unterschätzt. Korrekterweise müsste eigentlich noch die Unsicherheit bei der Wahl

des Glättungsparameters und den Kovariablen für das zur Schätzung der Gewichte (5.5) notwendige GAM berücksichtigt werden.

- Die Verwendung von einfachen, nicht-multiplen Imputationen führt in der Regel zu sehr guten Schätzungen für alle FMA- und FMS-Schätzer. Häufig liefert die Amelia-Methode, also die Verwendung des *R*-Pakets „Amelia II“, sehr gute und stabile Ergebnisse. In einigen Situationen, etwa den Experimenten 3 und 7 aus Abschnitt 6.1 und den Experimenten 4,5 und 6 aus Abschnitt 6.2, kann jedoch auch die Verwendung der GAMRI-, GLMRI- bzw. kNN-Methodik zu den besten Resultaten führen. Prinzipiell kann nicht konstatiert werden, dass bei Gebrauch des in Abschnitt 5.1.2 vorgestellten verallgemeinerten Regressionsimputationsalgorithmus die Verwendung von generalisierten additiven Modellen (GAMRI) der Verwendung simpler generalisierter linearer Regressionsmodelle (GLMRI) vorzuziehen ist. Der Erfolg dieser beiden Methoden variiert sehr stark mit dem gewählten Setting. Generell wird bei allen vier betrachteten Imputationsmethoden die Varianz unterschätzt, da die Imputationsunsicherheit nicht erfasst wird.
- Die Verwendung von korrekten multiplen Imputationen unter Verwendung des Amelia II-Pakets der statistischen Software *R* führt – wie in Abschnitt 6.3 diskutiert – zu guten Schätzungen, insbesondere auch in Bezug auf die Varianz, die speziell für den FMA-Akaike-Schätzer nicht unterschätzt wird. Wirkt der Fehlendmechanismus auf eine Variable, die einen komplexeren Effekt, etwa einen Interaktionseffekt, mitgestaltet und ist die Imputationsunsicherheit groß, so wird diese Unsicherheit explizit mitmodelliert und kann zu einer Überschätzung der Varianz führen. Dies verdeutlicht noch einmal die großen Herausforderungen, die in der Modellbildung unter Berücksichtigung fehlender Daten liegen. Ferner werden in den betrachteten Situationen die Standardfehler besser geschätzt, wenn zuerst die Modellselektionsunsicherheit und dann die Imputationsunsicherheit berücksichtigt wird und nicht umgekehrt.

Die zu Beginn dieses Kapitel aufgeworfenen Fragen lassen sich unter Berücksichtigung dieser Erkenntnisse wie folgt beantworten:

1. Modellmittelungsverfahren sind der Modellselektion nicht „generell“ überlegen, dies zeigen insbesondere die Ergebnisse der Mallows-Model-Averaging-Methodologie von Hansen (2007). Bei Betrachtung zweier spezifischer Verfahren, wie in diesem Kapitel etwa der FMA-Akaike- und FMS-AIC-Schätzer, kann jedoch konsta-

tiert werden, dass die Modellmittelung in der Regel deutlich bessere Schätzungen liefert als die Modellselektion.

2. Schätzer die Optimalitätseigenschaften besitzen sind solchen die eher pragmatisch motiviert sind nicht zwingenderweise überlegen. Dies folgt eindeutig aus dem Vergleich des FMA-Akaike- und FMA-Hansen-Schätzers.
3. Um auf den Sachverhalt fehlender Werte einzugehen, bieten sich viele Möglichkeiten an. In der Regel versprechen Imputationsverfahren die besten Ergebnisse. Einfache Imputationen liefern respektable Punktschätzungen, unterschätzen jedoch die zugehörige Varianz; Multiple Imputationen führen ebenfalls zu guten Punktschätzungen, können in einigen Situationen jedoch zu einer Überschätzung der entsprechenden Varianz führen. Der finale Erfolg der einzelnen Imputationsmethoden hängt von der gewählten Situation ab.

7. Anwendungsbeispiele

Die folgenden drei Abschnitte illustrieren die Verwendung der vorgestellten FMS- und FMA-Schätzer unter spezieller Berücksichtigung der Problematik fehlender Daten. Abschnitt 7.1 betrachtet dabei das lineare Regressionsmodell, Abschnitt 7.2 das logistische Regressionsmodell und Abschnitt 7.3 diskutiert die Chancen und Risiken der vorgestellten Methoden im Kontext der Faktorenanalyse.

7.1 Phasenangepasste Führung von Wachstumsunternehmen

Die folgenden beiden Abschnitte betrachten und erweitern die Analysen von Klaußner (2007), der den Erfolg von Unternehmen in der Life-Science-Branche abhängig vom Führungsverhalten und abhängig von der Unternehmenslebensphase untersucht. Die Methoden, die dabei von Klaußner (2007) verwendet werden, sind vorwiegend deskriptiver Natur, enthalten aber auch lineare Regressionsanalysen, die unter Berücksichtigung der Erkenntnisse der ersten sechs Kapitel überprüft, adjustiert und in den Kontext fehlender Daten gebracht werden. Die verwendeten Daten stammen dabei aus $n = 101$ Unternehmen und messen die interessierenden Größen wie folgt:

- Als „Erfolg“ wird jede Art von Führungserfolg verstanden: Dies meint explizit 1) die *Zufriedenheit* der Mitarbeiter mit ihrer Führung, 2) die *Leistungsbereitschaft* eines jeden Einzelnen, die über das grundsätzliche Maß einer Führer-Geführten-Beziehung hinausgeht und 3) die *Effektivität* der Führung, eine Maßzahl für die wirksame Erfüllung der Aufgaben durch den Geführten. Diese drei Größen werden über den Multifactor Leadership Questionnaire (MLQ, vgl. Bass, Avolio und Active (2003)) anhand eines Scores bestimmt. Der Score ist dabei der Mittelwert von vier Fragen auf einer Skala zwischen null und vier. Abschnitt 7.1.1 betrachtet eine lineare Regressionsanalyse für die Zufriedenheit, Abschnitt 7.1.2 eine lineare Regressionsanalyse für die Effektivität; die Leistungsbereitschaft als Indikator für Erfolg wird in dieser Arbeit nicht näher untersucht.

- Das „Führungsverhalten“ wird durch eine Einteilung von Bass und Steyrer (1995) charakterisiert, die versuchen das gesamte bekannte Führungsspektrum über neun verschiedene Führungsstile abzubilden. Die Intensität jedes einzelnen Führungsstils wird dabei erneut über einen Score zwischen null und vier mit Hilfe des MLQ erfasst. Bass und Steyrer (1995) ordnen dabei – wie in der einschlägigen Literatur üblich – die entsprechenden Führungsstile sowohl transaktionalen als auch transformationalen Führungsverhalten zu:
 - Transaktionales Führungsverhalten beschreibt dabei eine einfache Austauschbeziehung zwischen Führungskraft und Geführtem; dies bedeutet, dass das Verhältnis zwischen beiden Seiten als eine Transaktion aufgefasst wird: Der Geführte leistet genau das, was von ihm erwartet wird und bekommt dafür als Gegenleistung von der Führungspersönlichkeit das, was vorher vereinbart wurde, beispielsweise ein höheres Gehalt oder eine Beförderung. Ein solcher Führungsstil beinhaltet Verhaltensweisen, die mit dem Multifactor Leadership Questionnaire über die Variablen *Contingent Reward (CR)*, *Management by Exception Active (MEA)* und *Management by Exception Passive (MEP)* abgebildet werden. Für eine detaillierte Motivation und eine ausführliche Unterscheidung dieser Größen wird auf Klaußner (2007, Seite 88ff.) sowie die folgenden Analysen verwiesen.
 - Transformationale Führung meint jene Art von Führung, die es vermag, Leistungen freizusetzen, die über dem liegen, was von Mitarbeitern unter „normalen Bedingungen“ erwartet wird. Dies beinhaltet insbesondere charismatische Führung und Inspiration, intellektuelle Stimulation und eine individualisierte Betrachtung der Mitarbeiter. Der MLQ bildet dieses Führungsverhalten über die Variablen *Idealized Influence Attributed (IIA)*, *Idealized Influence Behavioral (IIB)*, *Inspirational Motivation (IM)*, *Intellectual Stimulation (IS)* und *Individual Consideration (IC)* ab.
 - Betrachtet wird ferner auch ein *Laisser-faire-Führungsstil*, der als Sonderfall zu verstehen ist; er ist weder transaktional noch transformational und zeichnet sich durch ein Höchstmaß an Ineffektivität aus, da bei zu bewältigenden Aufgaben nicht entschieden gehandelt wird und dadurch möglicherweise erst Probleme entstehen.

- Um die „Lebensphase“ eines Unternehmens zu charakterisieren, wird das *Lebenszyklus-Modell* von Pümpin und Prange (1991) betrachtet: Die Autoren gehen davon aus, dass jedes Unternehmen die Phasen Entstehung (=1), Wachstum (=2), Reife (=3) und Niedergang (=4) durchlebt – eventuell wiederholen sich diese Phasen nach vielen Jahren und auf die Niedergangsphase erfolgt eine Bereinigung und neues Entstehen und Wachstum. Jedes der 101 untersuchten Unternehmen wird auf Basis eines Fragebogens in eine dieser Phasen eingeteilt.

Ziel der Untersuchung ist es, den Erfolg eines Unternehmens durch ein phasenspezifisches Führungsverhalten zu charakterisieren. Dies beinhaltet die folgenden Fragestellungen: Welches Führungsverhalten steigert bzw. vermindert den (Führungs-)Erfolg in großem Ausmaß? Sind in verschiedenen Unternehmenslebensphasen verschiedene Führungsstile notwendig?

7.1.1 Analyse der Zufriedenheit

In diesem Abschnitt wird der Erfolgsindikator Zufriedenheit betrachtet. Es stellt sich die Frage, durch welches Führungsverhalten sie besonders gesteigert bzw. vermindert werden kann und ob dies abhängig von der Unternehmenslebensphase ist oder nicht. Tabelle 7.1.1 zeigt die Korrelation nach Spearman zwischen der Zufriedenheit und allen interessierenden Größen. Es fällt sofort auf, dass die transformationalen Führungsstile

	IIA	IIB	IM	IS	IC	CR	MEA	MEP	L	Phase
Zufriedenheit	0.65	0.40	0.43	0.53	0.67	0.55	0.01	-0.37	-0.37	-0.45

Tab. 7.1: Korrelation nach Spearman zwischen der Zufriedenheit und den potentiellen Einflussgrößen

IIA, IIB, IM, IS und IC allesamt positiv mit der Zufriedenheit korrelieren. Transaktionales Führungsverhalten, erfasst durch CR, MEA und MEP, wirkt positiv, neutral sowie negativ auf den Erfolg, was dafür spricht, dass diese Größen unterschiedlich erfasst und interpretiert werden müssen. Ein Laissez-faire-Führungsstil korreliert – wie zu erwarten – negativ mit der Zufriedenheit. Die negative Korrelation für die Unternehmenslebensphase impliziert, dass für spätere Phasen (Reife, Niedergang) eine geringere Zufriedenheit mit dem Führungspersonal beobachtet werden kann.

Um nun die relevanten Variablen zu klassifizieren und die Stärke ihres Einflusses zu quantifizieren, wird eine lineare Regressionsanalyse mit der Zufriedenheit als Response und allen oben angegebenen Größen als potentiellen Kovariablen durchgeführt; die Unternehmenslebensphase wird dabei dummykodiert mit Phase 1 (=Entstehung) als Referenz. Um die vorgestellten Methoden zur Berücksichtigung fehlender Daten zu veranschaulichen, wird ein MCAR-Fehlendmechanismus eingeführt²⁰; für alle Kovariablen wird die konstante Fehlwahrscheinlichkeit

$$\pi_{\text{IIA}} = \dots = \pi_{\text{L}} = \pi_{\text{Phase}} = 0.1$$

angenommen. Dies resultiert aufgrund der vielen betrachteten Variablen zu 68.32% an Beobachtungen, bei denen mindestens ein Wert fehlt. Eine Complete Case Analyse kann also nur auf etwas mehr als 30% der Fälle zurückgreifen. Die Kandidatenmodelle sind dieselben wie bei Klaußner (2007, Seite 235), der aufgrund verschiedener Selektionsprozeduren und einigen ad-hoc Betrachtungen sieben plausible Modelle vorstellt, deren AIC-Werte jeweils sehr nahe beisammen liegen²¹:

$$\begin{aligned} M_1 : y &= \alpha + \gamma_1 \text{IIA} + \gamma_2 \text{IIB} + \gamma_3 \text{IM} + \gamma_4 \text{IS} + \gamma_5 \text{IC} + \gamma_6 \text{CR} + \gamma_7 \text{MEA} + \gamma_8 \text{MEP} \\ &\quad + \gamma_9 \text{L} + \gamma_{10} \text{P2} + \gamma_{11} \text{P3} + \gamma_{12} \text{P4} + \gamma_{13} \text{LP2} + \gamma_{14} \text{LP3} + \gamma_{15} \text{LP4}, \\ M_2 : y &= \alpha + \gamma_1 \text{IIA} + \gamma_5 \text{IC} + \gamma_7 \text{MEA} + \gamma_9 \text{L} + \gamma_{10} \text{P2} + \gamma_{11} \text{P3} + \gamma_{12} \text{P4} + \gamma_{13} \text{LP2} \\ &\quad + \gamma_{14} \text{LP3} + \gamma_{15} \text{LP4}, \\ M_3 : y &= \alpha + \gamma_1 \text{IIA} + \gamma_5 \text{IC} + \gamma_9 \text{L} + \gamma_{10} \text{P2} + \gamma_{11} \text{P3} + \gamma_{12} \text{P4} + \gamma_{13} \text{LP2} \\ &\quad + \gamma_{14} \text{LP3} + \gamma_{15} \text{LP4}, \end{aligned}$$

²⁰ Die Motivation für dieses Anwendungsbeispiel Daten künstlich zu verwerfen, liegt in der von Klaußner (2007, Seite 184 ff.) angesprochenen Problematik der Datenerhebung. Ursprünglich angedacht war eine günstige und schnelle Online-Befragung. Aufgrund der geringen Rücklaufquote und einer Vielzahl fehlender Werte sah sich der Autor dazu gezwungen, durch den Besuch mehrerer Fachmessen die Umfragen persönlich durchzuführen. Da sich ein solches Prozedere nicht nur als zeit-, sondern auch als kostenintensiv darstellt, ist die Sensitivität der von Klaußner (2007) gefundenen Resultate bezüglich fehlender Daten auch für zukünftige, eventuell zeitlich eingeschränkte Forschungsarbeiten relevant. Ein MCAR-Fehlendmechanismus dient dabei einer einfachen Illustration der Problematik; MAR- und MNAR-Fehlendmechanismen werden im nächsten Abschnitt 7.1.2 betrachtet und motiviert.

²¹ Für fünf der sieben betrachteten Modelle ist die Differenz der entsprechenden AIC-Werte zum geringsten AIC-Wert kleiner als 5: $\text{AIC}(M_2)=174.38$, $\text{AIC}(M_3)=173.28$, $\text{AIC}(M_4)=175.19$, $\text{AIC}(M_5)=175.93$, $\text{AIC}(M_6)=178.21$. Die AIC-Werte unterschieden sich geringfügig von denen von Klaußner (2007, Abbildung 87), da dieser teilweise gerundete Scores für die Analyse verwendet.

$$M_4 : y = \alpha + \gamma_1 IIA + \gamma_5 IC + \gamma_7 MEA + \gamma_9 L + \gamma_{10} P2 + \gamma_{11} P3 + \gamma_{12} P4 ,$$

$$M_5 : y = \alpha + \gamma_1 IIA + \gamma_5 IC + \gamma_9 L + \gamma_{10} P2 + \gamma_{11} P3 + \gamma_{12} P4 ,$$

$$M_6 : y = \alpha + \gamma_1 IIA + \gamma_5 IC + \gamma_9 L ,$$

$$M_7 : y = \alpha + \gamma_1 IIA + \gamma_5 IC .$$

Diese Auswahl deutet darauf hin, dass die transformationalen Führungsstile IIA und IC, der transaktionale Führungsstil MEA und ein Laisser-faire Verhalten (teilweise in Interaktion mit der Unternehmenslebensphase) von besonderem Interesse sind. Um die Daten zu analysieren und die vorgestellten Methoden geeignet zu illustrieren, werden dieselben 20 Schätzer wie in Tabelle 6.1 betrachtet und ferner der FMA-Akaike-, FMA-Hansen- bzw. FMS-AIC-Schätzer nach multipler Imputation gemäß Tabelle 6.7. Es ist klar, dass alle FMA-Hansen-Schätzer das Modell M_1 und die entsprechenden 15 Submodelle betrachten. Die Ergebnisse der Analyse befinden sich in Tabelle 7.2.

Resultate

Es lassen sich folgende Resultate bezüglich der inhaltlichen Interpretation, der drei betrachteten Schätzer und der verschiedenen Methoden zur Berücksichtigung der fehlenden Werte konstatieren:

- Betrachtet man die Schätzungen der Regressionsparameter sowie deren geschätzte Standardfehler für die Originaldaten, so ist zu erkennen, dass alle Schätzer ein IIA-Führungsverhalten als besonders relevant für die Zufriedenheit einstufen; die Parameterschätzungen sind dabei stets mindestens dreimal so groß wie der zugehörige Standardfehler. Insbesondere der FMA-Akaike-, aber auch der FMS-AIC-Schätzer ordnet darüber hinaus auch IC eine größere Bedeutung zu. Dabei wirken die Führungsstile IIA und IC positiv auf die Zufriedenheit. Dies war aufgrund der betrachteten Korrelationen zu erwarten. Alle anderen Größen scheinen keinen relevanten Effekt auf die Zufriedenheit zu besitzen.
- Generell erfassen die Schätzer auf Basis der betrachteten Imputationsansätze die Struktur ähnlich wie die entsprechenden Schätzer auf Basis der Originaldaten. Der Hauptunterschied liegt im Zusammenspiel von Phase und Laisser-faire: Während die Verwendung multipler Imputationen (MI) bzw. des GAMRI-Algorithmus ähnlich den Originaldaten sowohl Phase als auch Laisser-faire als wenig relevant einstufen, so gesteht die kNN- bzw. Amelia-Methodik Laisser-faire an sich und das

	I.	IIA	IIB	IM	IS	IC	CR	MEA									
FMA-Akaike-Schätzer																	
1) Original	0.91	(0.43)	0.38	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.37	(0.10)	0.00	(0.00)	-0.04	(0.07)	...
2) CC	1.68	(2.22)	0.08	(0.31)	-0.07	(0.30)	-0.24	(0.35)	-0.04	(0.39)	0.32	(0.31)	0.51	(0.44)	-0.24	(0.23)	...
3) GAMRI	0.66	(0.44)	0.38	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.37	(0.10)	0.00	(0.00)	-0.01	(0.03)	...
GLMRI	0.47	(0.48)	0.32	(0.11)	-0.01	(0.03)	0.00	(0.02)	0.02	(0.04)	0.45	(0.11)	0.05	(0.09)	-0.07	(0.09)	...
kNN	1.23	(0.47)	0.31	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.42	(0.10)	0.00	(0.01)	-0.09	(0.10)	...
Amelia	1.49	(0.46)	0.31	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.41	(0.10)	0.00	(0.01)	-0.13	(0.10)	...
MI	1.14	(0.63)	0.36	(0.10)	0.00	(0.01)	0.01	(0.03)	0.00	(0.02)	0.39	(0.11)	0.01	(0.05)	-0.08	(0.10)	...
4) AIC _W	1.95	(1.31)	0.04	(0.26)	-0.07	(0.18)	-0.07	(0.21)	-0.04	(0.21)	0.42	(0.26)	0.24	(0.36)	-0.24	(0.19)	...
FMA-Hansen-Schätzer																	
1) Original	0.71	(0.59)	0.38	(0.14)	-0.07	(0.09)	0.02	(0.10)	0.11	(0.13)	0.26	(0.16)	0.07	(0.11)	-0.06	(0.08)	...
2) CC	1.74	(0.72)	0.33	(0.26)	-0.08	(0.16)	-0.08	(0.18)	-0.04	(0.15)	0.12	(0.20)	0.17	(0.25)	-0.01	(0.02)	...
3) GAMRI	0.55	(0.60)	0.36	(0.15)	-0.04	(0.09)	0.06	(0.12)	0.07	(0.13)	0.29	(0.17)	0.04	(0.08)	-0.01	(0.04)	...
GLMRI	0.44	(0.60)	0.28	(0.14)	-0.08	(0.10)	0.01	(0.12)	0.11	(0.13)	0.33	(0.15)	0.22	(0.17)	-0.11	(0.10)	...
kNN	0.95	(0.60)	0.28	(0.14)	-0.08	(0.09)	0.04	(0.10)	0.08	(0.13)	0.29	(0.14)	0.17	(0.15)	-0.11	(0.10)	...
Amelia	1.23	(0.53)	0.29	(0.14)	-0.04	(0.09)	0.03	(0.10)	0.03	(0.12)	0.28	(0.15)	0.19	(0.15)	-0.15	(0.11)	...
MI	1.08	(0.79)	0.34	(0.15)	-0.03	(0.10)	-0.03	(0.16)	0.04	(0.15)	0.26	(0.16)	0.19	(0.18)	-0.13	(0.12)	...
FMS-AIC-Schätzer																	
1) Original	0.73	(0.38)	0.39	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.36	(0.10)	0.00	(0.00)	0.00	(0.00)	...
2) CC	1.68	(2.22)	0.08	(0.31)	-0.07	(0.30)	-0.24	(0.35)	-0.04	(0.39)	0.32	(0.31)	0.51	(0.44)	-0.24	(0.23)	...
3) GAMRI	0.59	(0.41)	0.38	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.36	(0.10)	0.00	(0.00)	0.00	(0.00)	...
GLMRI	0.30	(0.38)	0.36	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.45	(0.10)	0.00	(0.00)	0.00	(0.00)	...
kNN	1.42	(0.42)	0.29	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.42	(0.09)	0.00	(0.00)	-0.15	(0.08)	...
Amelia	1.62	(0.39)	0.29	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.42	(0.10)	0.00	(0.00)	-0.18	(0.08)	...
MI	1.25	(0.61)	0.34	(0.10)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.39	(0.10)	0.00	(0.00)	-0.10	(0.12)	...
4) AIC _W	1.76	(1.83)	0.06	(0.29)	-0.14	(0.29)	-0.14	(0.34)	-0.09	(0.37)	0.36	(0.30)	0.48	(0.41)	-0.24	(0.22)	...

Tab. 7.2: Schätzung der Regressionsparameter (Teil I), in Klammern die zugehörigen Standardfehler gemäß (5.24)

		MEP		L		P2		P3		P4		LP2		LP3		LP4	
FMA-Akaike-Schätzer																	
1) Original	...	0.00	(0.00)	-0.07	(0.23)	0.27	(0.28)	-0.24	(0.46)	0.06	(0.43)	-0.14	(0.23)	0.11	(0.27)	-0.36	(0.34)
2) CC	...	-0.02	(0.21)	0.43	(1.23)	0.41	(0.97)	0.36	(1.46)	0.05	(1.26)	-0.77	(1.24)	-0.45	(1.38)	-0.68	(1.30)
3) GAMRI	...	0.00	(0.00)	0.04	(0.26)	0.41	(0.32)	-0.16	(0.48)	0.38	(0.46)	-0.19	(0.27)	0.06	(0.31)	-0.56	(0.35)
GLMRI	...	0.00	(0.02)	0.26	(0.25)	0.52	(0.31)	0.00	(0.46)	0.47	(0.43)	-0.46	(0.28)	-0.17	(0.32)	-0.70	(0.34)
kNN	...	0.00	(0.00)	-0.28	(0.12)	0.12	(0.21)	0.05	(0.25)	-0.32	(0.29)	0.01	(0.05)	0.02	(0.07)	-0.02	(0.06)
Amelia	...	0.00	(0.00)	-0.29	(0.12)	-0.03	(0.22)	-0.13	(0.30)	-0.46	(0.30)	0.02	(0.08)	0.05	(0.12)	-0.03	(0.08)
MI	...	0.00	(0.01)	-0.26	(0.25)	0.10	(0.35)	-0.13	(0.45)	-0.17	(0.44)	0.02	(0.23)	0.12	(0.28)	-0.09	(0.22)
4) AIC _W	...	-0.01	(0.10)	0.02	(0.80)	0.11	(0.28)	0.07	(0.31)	-0.06	(0.19)	-0.20	(0.40)	-0.03	(0.27)	-0.07	(0.20)
FMA-Hansen-Schätzer																	
1) Original	...	0.01	(0.05)	0.03	(0.16)	0.30	(0.28)	-0.12	(0.39)	-0.02	(0.47)	-0.14	(0.23)	0.01	(0.19)	-0.30	(0.36)
2) CC	...	0.00	(0.00)	0.00	(0.00)	-0.03	(0.36)	-0.03	(0.41)	-0.52	(0.49)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
3) GAMRI	...	-0.01	(0.04)	0.04	(0.14)	0.34	(0.28)	-0.07	(0.38)	0.06	(0.50)	-0.09	(0.19)	0.02	(0.17)	-0.28	(0.37)
GLMRI	...	0.01	(0.06)	0.16	(0.23)	0.40	(0.34)	0.07	(0.41)	0.19	(0.52)	-0.32	(0.33)	-0.16	(0.28)	-0.45	(0.42)
kNN	...	0.00	(0.06)	-0.20	(0.15)	0.16	(0.21)	0.10	(0.23)	-0.33	(0.27)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Amelia	...	0.01	(0.07)	-0.22	(0.14)	0.03	(0.22)	-0.04	(0.23)	-0.47	(0.26)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
MI	...	0.01	(0.06)	-0.20	(0.27)	0.12	(0.36)	-0.04	(0.39)	-0.23	(0.41)	0.01	(0.20)	0.06	(0.22)	-0.06	(0.18)
FMS-AIC-Schätzer																	
1) Original	...	0.00	(0.00)	0.01	(0.23)	0.36	(0.29)	-0.38	(0.49)	0.29	(0.36)	-0.20	(0.26)	0.17	(0.34)	-0.55	(0.28)
2) CC	...	-0.02	(0.21)	0.43	(1.23)	0.41	(0.97)	0.36	(1.46)	0.05	(1.26)	-0.77	(1.24)	-0.45	(1.38)	-0.68	(1.30)
3) GAMRI	...	0.00	(0.00)	0.09	(0.26)	0.46	(0.31)	-0.21	(0.51)	0.51	(0.39)	-0.23	(0.28)	0.08	(0.36)	-0.66	(0.29)
GLMRI	...	0.00	(0.00)	0.34	(0.22)	0.62	(0.27)	-0.06	(0.49)	0.64	(0.37)	-0.52	(0.25)	-0.16	(0.33)	-0.83	(0.27)
kNN	...	0.00	(0.00)	-0.29	(0.09)	0.10	(0.19)	0.07	(0.22)	-0.42	(0.24)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Amelia	...	0.00	(0.00)	-0.29	(0.08)	-0.04	(0.19)	-0.07	(0.22)	-0.55	(0.24)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
MI	...	0.00	(0.00)	-0.29	(0.20)	0.06	(0.32)	-0.10	(0.48)	-0.29	(0.37)	0.07	(0.20)	0.11	(0.32)	-0.02	(0.14)
4) AIC _W	...	-0.02	(0.18)	0.33	(1.00)	0.16	(0.34)	0.08	(0.38)	-0.03	(0.23)	-0.34	(0.51)	-0.12	(0.38)	-0.14	(0.25)

Tab. 7.2: Schätzung der Regressionsparameter (Teil II), in Klammern die zugehörigen Standardfehler gemäß (5.24)

8. Résumé

Die in dieser Arbeit gewonnenen Erkenntnisse sind äußerst vielfältig und in vielerlei Hinsicht aufschlussreich für die Beurteilung des Erfolgs unterschiedlicher Strategien zur Berücksichtigung fehlender Werte im Kontext von Modellwahl und Modellmittelung. Auch wenn Resultate auf Basis von Monte-Carlo-Simulationen und Datenbeispielen immer einer gewissen Beschränkung unterliegen und nur mit Vorsicht auf Situationen außerhalb der betrachteten übertragen werden können, so ergeben sich doch Verhaltensweisen der verschiedenen Schätzer, die relativ durchgängig beobachtet und erklärt werden können und insofern auch einige typische Charakteristika aufweisen:

- Ungeachtet der Problematik fehlender Daten lässt sich festhalten, dass eine einfache, kriteriums-basierte Modellmittlungsstrategie, wie die Verwendung exponentieller AIC-Gewichte und damit des FMA-Akaike-Schätzers, innerhalb einer linearen oder logistischen Regressionsanalyse zu sehr guten Ergebnissen führt. In der Regel erhält man bei einem solchen Vorgehen bessere Punktschätzungen als bei den entsprechenden Selektionsschätzern. Die zugehörigen Varianzschätzungen sind in der Regel ebenfalls von besserer Qualität. Wie vielfach diskutiert, ist die Konstruktion von FMA-Schätzern unter Optimalitätseigenschaften durchaus erstrebenswert, auch unter Berücksichtigung des bisher relativ wenig einheitlichen Vorgehens innerhalb der frequentistischen Modellmittelung. Der in dieser Arbeit betrachtete MMA-Schätzer von Hansen (2007) führt jedoch in vielen Fällen zu keiner Verbesserung im Vergleich zu einer simplen Modellselektion, etwa unter Verwendung des AIC. Dies mag häufig, jedoch nicht immer, an seiner restriktiven Auswahl der Kandidatenmodelle liegen.
- Werden die fehlenden Werte einer Datenmatrix schlicht verworfen, also nur die vollständigen Fälle (CC) zur Analyse verwendet, so führt dies in den betrachteten Regressionsbeispielen, unabhängig von dem gewählten Selektions- oder Mittelungsverfahren, bei einem MAR-Fehlendmechanismus zu minderwertigen, insgesamt relativ unbefriedigenden Schätzungen. Dies war zu erwarten. Die vorliegende Arbeit

zeigt jedoch an vielen Stellen, dass bei einer CC-Strategie in der Regel immerhin die entsprechenden Varianzschätzungen von passabler Qualität sind.

- Die Verwendung des für fehlende Daten adjustierten AIC_W , entweder direkt zur Selektion des Modells oder alternativ zur Konstruktion von Modellmittelungsgewichten, führt insbesondere im Kontext des linearen Modells meist zu besseren Punktschätzungen als eine Complete Case Analyse, aber auch zu etwas schlechteren Ergebnissen als die Imputationsverfahren. In den betrachteten Situationen der logistischen Regressionsanalyse kann die AIC_W -Methodologie jedoch häufig keinen zusätzlichen Gewinn erbringen. Berechnet man die Varianz des FMA- AIC_W -Schätzers unter Berücksichtigung der Modellselektionsunsicherheit wie in (5.19) angeführt, so ist eine deutliche Unterschätzung der Varianz zu erkennen. Wie in dieser Arbeit bereits vielfach erläutert, müsste korrekterweise eigentlich noch die Unsicherheit bei der Wahl des Glättungsparameters und den Kovariablen für das zur Schätzung der Gewichte (5.5) notwendige GAM berücksichtigt werden. Dies zeigt, dass die Konzepte des *inverse probability weighting* innerhalb des Themenkomplexes von Modellwahl und -mittelung zu einer weiteren Unsicherheitskomponente führen, deren Modellierung gegebenenfalls sehr aufwändig sein kann.
- Die fehlenden Werte einer Datenmatrix zu ersetzen und anschließend die entsprechenden FMS- und FMA-Schätzer zu berechnen, kann als relativ weitläufige Strategie angesehen werden, da sowohl die Imputations- als auch die Selektions- bzw. Mittelungsverfahren frei gewählt werden können, was insbesondere in einem so schwer einzugrenzenden Feld wie der Modellwahl einen großen Vorteil darstellt. Insgesamt lassen sich in den angeführten Beispielen sehr stabile und gute Punktschätzungen erkennen, wobei deren Varianz bei Verwendung nicht-multipler Imputationen aufgrund der nicht modellierten Imputationsunsicherheit prinzipiell etwas unterschätzt wird. Die in dieser Arbeit verwendeten Imputationsverfahren stehen stellvertretend für eine Vielzahl an Methoden unterschiedlichster Konzeption. Häufig liefert dabei die Amelia-Methode, also die Verwendung des *R*-Pakets „Amelia II“, sehr gute und stabile Ergebnisse. In einigen Situationen, etwa bei Betrachtung komplexerer Abhängigkeitsstrukturen oder zufälligen Effekten, kann jedoch auch die Verwendung der GAMRI-, GLMRI- bzw. kNN-Methodik zu den besten Resultaten führen. Prinzipiell kann nicht konstatiert werden, dass bei Gebrauch des in Abschnitt 5.1.2 vorgestellten verallgemeinerten, rekursiven Regressionsimputationsalgorithmus die Verwendung von generalisierten additiven Model-

len (GAMRI) der Verwendung simpler generalisierter linearer Regressionsmodelle (GLMRI) vorzuziehen ist. Der Erfolg dieser beiden Methoden variiert mit dem gewählten Setting.

- Die Verwendung von korrekten multiplen Imputationen unter Verwendung des Amelia II-Pakets der statistischen Software *R* kann prinzipiell zu recht guten und realistischen Schätzungen führen, insbesondere in Bezug auf den FMA-Akaike-Schätzer. Selbst dort kann jedoch unter gewissen Umständen ein interessantes Artefakt beobachtet werden: Wirkt der Fehlendmechanismus auf eine Variable, die einen komplexeren Effekt, etwa einen Interaktionseffekt, mitgestaltet und ist die Imputationsunsicherheit groß, so wird diese Unsicherheit wie gewünscht explizit mitmodelliert und kann zu einer Überschätzung der Varianz führen. Auch wenn nach einer Reflexion über diese Methodik ein solches Verhalten des Schätzers als plausibel erscheint, so ist dies bei der Interpretation konkreter Anwendungsbeispiele zu beachten.

Es hat sich gezeigt, dass die vorgestellten Konzepte zur Berücksichtigung fehlender Daten im Kontext von Modellselektion und Modellmittelung für lineare und logistische Regressionsanalysen einfach umzusetzen sind und in den gewählten Anwendungsbeispielen in Anbetracht ihrer oben angeführten Charakteristika zu entsprechend guten bzw. weniger guten Resultaten führen, auf jeden Fall aber gut interpretierbar sind. Sehr einfache Ansätze (kriteriums-basierte Modellmittelungsgewichte, Imputationen aus vorhandener Software) scheinen dabei oft zu bereits sehr guten Resultaten zu führen.

Etwas differenzierter gestaltet sich der Blick bei Verwendung der Verfahren in der Faktorenanalyse. Die vorliegende Arbeit macht deutlich, weshalb sich die aktuelle Literatur vorwiegend einfachster Regressionsanalysen bedient, um ihre Methoden zu illustrieren: Eine Übertragung der grundlegenden Konzepte ist zwar prinzipiell auf viele Typen statistischer Modellierung möglich und den Umständen entsprechend technisch unkompliziert; liegen der Interpretationsschwerpunkt und die generelle Konzipierung etwas weiter von den Prinzipien der Regressionsanalyse entfernt, so ergeben sich jedoch neue Schwierigkeiten, Unsicherheitskomponenten und gegebenenfalls auch ein geringerer Informationsgewinn. In der Faktorenanalyse wird dies besonders deutlich, da hier in der Regel die Punkt- und nicht die zugehörigen Varianzschätzungen im Vordergrund stehen und insofern die Modellierung zusätzlicher, durch die Modellselektionsunsicherheit hervorgerufener, Unsicherheit keinen wirklichen Gewinn erbringt. Auch ist die Anzahl der konkurrierenden Kandidatenmodelle im Vergleich zur Regressionsanalyse, wo In-

teraktionen, logarithmische und quadratische Effekte oft zu einer großen Auswahl an Modellierungsmöglichkeiten führen, sehr beschränkt, was den Erfolg ebenfalls limitiert.

Ausblick

Wie die Problematik fehlender Daten im Kontext von Modellmittlungsverfahren berücksichtigt werden kann und welche Chancen, aber auch Schwierigkeiten, die vorgestellten Schätzer bieten, hat die vorliegende Arbeit aufgezeigt. Viele Ideen sind sehr allgemein formuliert worden und eröffnen damit die Möglichkeit, jenseits einfacher linearer und logistischer Regressionsanalysen verwendet zu werden. Tatsächlich stellt sich die Frage, wie einfach die Konzepte auf komplexere Modelle übertragen werden können.

In den letzten Jahren haben einige erste Arbeiten aufgezeigt, welchen Erfolg neuere (frequentistische) Modellmittlungskonzepte in anderen Bereichen versprechen: Hjort und Claeskens (2006) verwenden kriteriums-basierte Gewichte unter Verwendung des Focused Information Criterion (FIC) von Claeskens und Hjort (2003), um einen FMA-Schätzer im Cox-Modell zu konstruieren, Hansen (2008b) überträgt einige Kerngedanken seines MMA-Schätzers (Hansen (2007)) auf den Kontext spezieller autoregressiver Prozesse und Zhang, Wan und Zhou (2010) verwenden die Ideen von Claeskens und Hjort (2003) und Hjort und Claeskens (2006) im Kontext des Tobit-Modells. Sowohl die „Mittlung nach Imputation“-Strategie als auch die Idee des *inverse probability weighting* könnten auch hier beim Vorhandensein fehlender Werte verwendet werden. Welchen Erfolg die in dieser Arbeit behandelten Imputationsmethoden jedoch versprechen und wie das Gewichtungskonzept konkret etwa auf das FIC angewendet werden kann, bedarf noch einer genaueren, möglicherweise simulationsgestützten Untersuchung.

Auch stellt sich die Frage, wie Modellmittlung und fehlende Daten, etwa im Bereich nonparametrischer Regression, kombiniert werden können. Die einfachste Möglichkeit besteht sicher darin, multiple Imputationen und exponentielle Gewichte auf Basis von Vorhersagefehlern zu verwenden, etwa mit Hilfe des generalisierten Kreuzvalidierungskriteriums, das für viele nonparametrische Ansätze – etwa bei additiven Modellen – von statistischer Software standardmäßig ausgegeben wird. Diskussionswürdig erscheint dagegen der Gegenstand der Mittlung: Soll über den gesamten Prädiktor gemittelt werden, ähnlich dem Vorgehen beim Bagging (Breiman (1996b))? Dies würde sicherlich einen Teil der Selektionsunsicherheit abdecken, für den Großteil der Verfahren im Bereich nonparametrischer Regression umsetzbar sein und möglicherweise auch zu besseren Vorhersagen führen. Damit kann aber zum einen keine Interpretation der einzelnen Effekte

und Variablen erfolgen und zum anderen wird die Unsicherheit bezüglich der Wahl des Glättungsparameters und der Variablen nicht ausreichend abgedeckt. Der erstgenannte Punkt kann sicher dahingehend berücksichtigt werden, dass nicht über den gesamten Prädiktor, sondern über die einzelnen Effekte oder bei Verwendung stückweiser Polynome sogar über die Parameterschätzungen gemittelt wird. Konkret kann die Entscheidung wohl von der bevorzugten Schätzmethodik und dem Ziel der Analyse abhängig gemacht werden. Der zweite Punkt lässt sich möglicherweise erfassen, indem doppelt kombiniert wird, was bedeutet, dass die Mittelung über einen Grid erfolgt, der sowohl eine Variablen- und Effektauswahl als auch mögliche Smoothing-Parameter enthält. Ob sich ein solch großer Aufwand überhaupt lohnt, ist aber sicherlich fraglich.

Insgesamt erscheinen auch unabhängig vom Auftreten fehlender Daten die Konzepte frequentistischer Modellmittelung im Gegensatz zur bayesianischen Herangehensweise noch nicht wirklich einheitlich und in sich konsistent. Unter welchem Paradigma die Modelle kombiniert werden sollen und welche Eigenschaften und Verteilung diese Schätzer besitzen, ist bisher nicht ausreichend geklärt. Speziell für die Konstruktion von Konfidenzintervallen ist dies jedoch besonders interessant. Hjort und Claeskens (2003) zeigen zwar, dass innerhalb eines lokalen Misspezifikationskonstrukts und unter gewissen Regularitätsvoraussetzungen die asymptotische Verteilung eines Modellmittelungsschätzers einer konvexen Kombination von Normalverteilungen entspricht, Leeb und Pötscher (2006b) zweifeln die Nützlichkeit dieses Resultats jedoch an, da die Konvergenz gegen die wahre Verteilungsfunktion nicht gleichmäßig, sondern nur punktweise in θ gilt und damit für jede fixe Stichprobengröße die wahren und geschätzten Verteilungen potentiell weit voneinander entfernt liegen können. Ob diese Aussagen tatsächlich von praktischer Relevanz sind und ob im Kontext fehlender Werte für die Berechnung von Konfidenzintervallen nicht die Verwendung von simplen Normalverteilungsquantilen, wie auch in Burnham und Anderson (2002) vorgeschlagen, zusammen mit den Standardfehlern, wie in Kapitel 5 beschrieben, eine ausreichende Approximation liefern, müssen zukünftige Arbeiten zeigen. Die Simulationsergebnisse aus Abschnitt 6.3 lassen zumindest erahnen, dass die Berechnung der Standardfehler gemäß (5.24) unter Verwendung korrekter multipler Imputationen für einen FMA-Akaike-Schätzer sehr gute Ergebnisse erzielen kann und die entsprechenden Konfidenzintervalle damit zumindest in einem einigermaßen sinnvollen und adäquaten Bereich liegen können.

Prinzipiell scheinen noch viele Erweiterungen der vorgeschlagenen Methodik möglich: So kann durchaus angedacht werden, die angeführten Korrekturverfahren für fehlende Daten im Bereich bayesianischer Modellmittelung zu untersuchen. Dort stellen sich

natürlich weiterhin die in den Abschnitten 3.4 und 4.1 aufgeworfenen, offenen Fragen. Um es kurz zu sagen: Wieviel Bayes soll sein? Weniger bis gar kein Bayes, indem die SBC-Approximation zur schnellen Berechnung des BMA-Schätzers verwendet wird und damit keine priori-Wahrscheinlichkeiten $p(M_k)$ spezifiziert und interpretiert werden müssen, oder mehr Bayes, indem mit Laplace-Approximationen genauere, aber auch aufwändigere Lösungen gesucht werden. Eine Verwendung von (ggf. multiplen) Imputationsmethoden ist hier immer möglich, auch ein Gewichtungsansatz könnte insbesondere bei Verwendung von exponentiellen SBC-Gewichten konstruiert werden.

Auch ist weiterhin offen, ob die multiplen Imputationen über den im Amelia II-Paket implementierten Bootstrap-Ansatz zu qualitativ ähnlichen Ergebnissen führen wie multiple Imputationen auf Basis eines streng bayesianischen Vorgehens, wie etwa unter Verwendung des IP-Algorithmus. Möglicherweise gibt es an dieser Stelle noch etwas Spielraum um die Varianzschätzungen nach (5.24) noch etwas stabiler zu machen und die Überschätzung der Standardfehler in komplexen Situation eher zu unterbinden. Die Analyse und die Konstruktion von Varianzschätzern scheint im Rahmen frequentistischer Modellmittelung allgemein noch ausbaufähig zu sein. Ein Großteil aktueller Artikel, so etwa von Yang (2003), Yuan und Yang (2005), Hansen (2007, 2008a,b, 2009), Magnus, Powell und Prüfer (2008), Liang et al. (2010), Hansen und Racine (2009) und Zhang, Wan und Zhou (2010) betrachtet vor allem die Qualität der Punktschätzungen, obwohl ein Hauptziel von Analysen post model selection sicherlich in einer geeigneten Erfassung der Variabilität dieser Schätzungen liegt.

Es bleibt festzuhalten, dass das Bewusstsein, geeignete Inferenzverfahren zur Berücksichtigung der Modellselektionsunsicherheit zu konstruieren, in den letzten Jahren deutlich geschärft wurde und viele Denkanstöße der Fachliteratur auch in den angewandten Wissenschaften immer häufiger verwendet werden. Ob dies auch im Kontext fehlender Daten gilt, werden zukünftige Arbeiten zeigen.