

Maik Hammerschmidt/Robert Wilken/Matthias Staat

# Methoden zur Lösung grundlegender Probleme der Datenqualität in DEA-basierten Effizienzanalysen



Maik  
Hammerschmidt



Robert Wilken

## ■ Schlüsselbegriffe

Bootstrapping-Verfahren; Data Envelopment Analysis; Datenqualität; Effizienzanalyse; Marketing Performance

## ■ Keywords

Data envelopment analysis (DEA); bootstrapping; data quality; modelling methodology; marketing performance

## Zusammenfassung

In diesem Beitrag werden grundlegende Aspekte der Datenqualität im Rahmen von Effizienzanalysen auf Basis der Data Envelopment Analysis (DEA) behandelt. Jede DEA benötigt die Spezifizierung einer Ursprungsdatenmatrix, deren Dimensionen durch die Analyseeinheiten sowie durch die verwendeten Merkmale (Input- und Outputvariablen) determiniert werden. Für die mit der Spezifizierung der Dimensionen zusammenhängenden Probleme stellen wir jeweils geeignete Lösungsansätze vor. Wir illustrieren diese anhand von drei empirischen Beispielen zur Effizienzanalyse von Vertragshändlern, Bankfilialen und Produktmärkten. So zeigen wir im Sinne eines Leitfadens, wie man in konkreten Anwendungssituationen mit den genannten Problemen der Datenqualität in DEA-basierten Effizienzanalysen verfahren sollte.

## Abstract

This paper addresses fundamental data quality issues in the context of efficiency measurement with Data Envelopment Analysis (DEA). Any DEA requires a data matrix to be specified, with units of analysis and efficiency criteria as dimensions. The choice of both units of analysis and efficiency criteria is crucial for the analysis, as it significantly influences the meaning of the results. We review the relevant literature that deals with these aspects of data quality in DEA. Furthermore, suitable methods are illustrated by empirical examples. Thus, the reader also gets to know how to use these methods in real-world applications.

## Autoren

Dr. Maik Hammerschmidt, Habilitand am Lehrstuhl für ABWL und Marketing II der Universität Mannheim, L 5, 1, 68131 Mannheim, E-Mail: maik.hammerschmidt@bwl.uni-mannheim.de, Tel.: +49 (0)621/181-1569, Fax: +49 (0)621/181-1571; Prof. Dr. Robert Wilken, Juniorprofessor für Internationales Marketing, ESCP-EAP European School of Management, Heubnerweg 6, 14059 Berlin, E-Mail: robert.wilken@escp-eap.de, Tel.: +49 (0)30/32007-167, Fax: +49 (0)30/32007-118; PD Dr. Matthias Staat, Privatdozent, Fakultät für Rechtswissenschaft und Volkswirtschaftslehre der Universität Mannheim, L 7, 3-5, 68131 Mannheim.  
Für wertvolle Hinweise danken wir den beiden anonymen Gutachtern sowie Dipl.-Vw. Gerrit Löber und Dipl.-Wirtsch.-Inf. Nico Shenawai.

## 1. Einleitung

Die Data Envelopment Analysis (im Folgenden: DEA) hat sich seit ihrer Einführung in die betriebswirtschaftliche Literatur durch Charnes/Cooper/Rhodes im Jahre 1978 zu einer weit verbreiteten und akzeptierten Methode der Effizienzmessung entwickelt. Mittlerweile existieren mehrere Tausend Publikationen in Form von Artikeln in wissenschaftlichen Zeitschriften, Büchern, Monografien und Arbeitspapieren zum Thema DEA; selbst ständig aktualisierte Bibliografien können dabei kaum den Anspruch auf Vollständigkeit erheben.<sup>1</sup> Diese Publikationen teilen sich in solche, die die DEA auf eine spezifische Fragestellung der Effizienzmessung anwenden, und andere mit eher theoretischem Anspruch, die das ursprüngliche Modell weiterentwickeln. Beiden Forschungsrichtungen ist gemein, dass sie zumindest indirekt mit Fragen der Gütebeurteilung konfrontiert sind: Im ersten Fall sind die Ergebnisse der Effizienzmessung in geeigneter Weise hinsichtlich ihrer Qualität zu bewerten. Im letzteren Fall geht es hauptsächlich darum, Modellformulierungen zu finden, die aus konzeptioneller Perspektive dem ursprünglichen Modell überlegen sind, eben weil sie eine der Annahmen jenes Modells adäquat abschwächen. Auch wenn aktuell nicht alle Probleme der Gütebeurteilung einer DEA zufriedenstellend gelöst worden sind, ist zumindest festzustellen, dass Ex-Post-Analysen – also solche Analysen, die zur *Ergebnis*beurteilung einer DEA eingesetzt werden – eine recht umfassende Behandlung in der Literatur erfahren haben.<sup>2</sup>

Im Gegensatz dazu findet man vergleichsweise wenig Literatur über Aspekte der Gütebeurteilung, die bereits *vor* der Durchführung der »eigentlichen« DEA behandelt werden müssen, also solche Aspekte, die sich nicht auf die Analyseergebnisse,

sondern auf die *Ursprungsdaten* beziehen. So beklagen Pedraja-Chaporro/Salinas-Jimenez/Smith (1999, S. 636) »... the absence of a convincing model-building methodology for the user of DEA.« Diese Aspekte der Datenqualität müssen im Rahmen von Ex-Ante-Analysen geklärt werden. Um dies näher erläutern zu können, wenden wir uns zunächst kurz dem einer DEA zugrunde liegenden Konzept zu.

Der ökonomische Effizienzbegriff hat seinen Ursprung im von Pareto (1897) formulierten wohlfahrtsökonomischen Prinzip und wurde von Koopmans (1951) auf die Produktionstheorie übertragen, womit die sog. Aktivitätsanalyse begründet war. Allgemein versteht man unter einer Aktivität die Realisierung eines Produktionsprozesses, der wiederum anhand von Inputs und Outputs beschrieben werden kann.<sup>3</sup> Zum effizienzorientierten Vergleich von Aktivitäten (oder auch: Produktionen<sup>4</sup>) werden als Kriterien die Inputs und Outputs herangezogen, wobei generell Inputs zu minimieren und Outputs zu maximieren sind. Eine Produktion bezeichnet man als Pareto-Koopmans-effizient genau dann, wenn sie nicht dominiert wird. Dominanz wiederum liegt vor, wenn eine DMU gegenüber jeder anderen bezüglich sämtlicher Kriterien mindestens genauso gut und bezüglich mindestens eines Kriteriums besser ist (man beachte hierbei die unterschiedlichen Optimierungsrichtungen bezüglich der Inputs und Outputs).

Die DEA ist ein nicht-parametrisches Verfahren zur methodischen Umsetzung dieses Effizienzkonzeptes. Analyseeinheiten der DEA stellen so genannte *decision making units* (DMUs) dar, beispielsweise Unternehmensfunktionen (wie Marketing oder Vertrieb) oder auch gesamte Unternehmen. Anhand der beobachteten Inputs und Outputs der DMUs konstruiert die DEA unter Annahme bestimmter Axiome eine so genannte Technologie-menge, die alle für möglich erachteten Produktionen enthält.<sup>5</sup> Die Quantifizierung der Effizienz einer DMU erfolgt durch Abstandsmessung des zugehörigen Input-Output-Vektors zum effizienten Rand der Technologie-menge, wobei Verfahren der Linearen Programmierung zum Einsatz gelangen.

Neben den grundlegenden Prinzipien einer DEA wird durch die bisherigen Ausführungen zugleich deutlich, dass sich die Ursprungsdaten einer DEA durch eine Matrix darstellen lassen, deren Zeilen die DMUs und deren Spalten die Kriterien – Inputs und Outputs – repräsentieren. Die oben angespro-

1 Die bereits 2002 erschienene Bibliografie von Tavares listet bspw. über 3.200 Veröffentlichungen von mehr als 1.600 Autoren aus 42 Ländern auf.

2 Zu einer Übersicht über derartige Probleme und Lösungsansätze vgl. beispielsweise den aktuellen Herausgeberband von Zhu/Cook (2007).

3 Vgl. Dyckhoff (1993). Die Begriffe »Aktivität« und »Produktion« werden dabei häufig synonym verwendet.

4 Trotz synonyme Verwendung der Begriffe »Aktivität« und »Produktion« ist auf die unterschiedliche theoretische Basis hinzuweisen; zu einer ausführlichen Darstellung der theoretischen Abgrenzung der beiden Begriffe sei auf Kleine (2002) verwiesen.

5 Vgl. Wilken (2007), S. 13 ff.

chenen Aspekte der Datenqualität innerhalb einer DEA betreffen nun die Festlegung der beiden »kanonischen« Dimensionen der Ursprungsdatenmatrix, konkret also die Wahl der Vergleichseinheiten und die Wahl der Inputs und Outputs, sowie zusätzlich das Verhältnis der beiden Dimensionen zueinander. Diesen zentralen Fragen werden wir im Folgenden nachgehen. Bezüglich der Festlegung der DMUs und der Kriterien sei angemerkt, dass wir die gängigen bisher verfügbaren Konzepte und Lösungsansätze diskutieren und jeweils auch anführen, welcher Ansatz aus konzeptioneller Perspektive überlegen ist. Es sei in diesem Zusammenhang allerdings darauf hingewiesen, dass es sich bei diesem Beitrag nicht um einen vollständigen Überblick handeln kann. Die Auswahl der hier vorgestellten Lösungsansätze erfolgt ausschließlich auf Basis von qualitativen Argumenten, die einzelne Ansätze anderen gegenüber vorteilhaft erscheinen lassen und somit eine besondere Behandlung »verdienen«.

Der Beitrag gliedert sich wie folgt: Zunächst stellen wir die Grundbegriffe und Notationen der DEA vor, insbesondere des DEA-Basismodells von Charnes/Cooper/Rhodes (Abschnitt 2). Diese Darstellung ist aufgrund der Verfügbarkeit in einschlägigen, auch deutschsprachigen Quellen bewusst knapp gehalten, jedoch vor allem aus Gründen der formalen Konsistenz innerhalb dieses Beitrages unerlässlich. In Abschnitt 3, dem Hauptteil, wenden wir uns den erwähnten drei grundsätzlichen Problembereichen der Datenqualität zu, wobei wir jeweils erstens das Problem erläutern, zweitens existierende Lösungsansätze vorstellen und drittens einen besonders brauchbaren Ansatz im Rahmen einer Kurzfallstudie unter Verwendung realer Datensätze illustrieren. Auf diese Weise wird dem Leser exemplarisch aufgezeigt, wie man mit den hier besprochenen Fragestellungen der Datenqualität bei DEA-basierten Effizienzanalysen im Rahmen einer konkreten Anwendungssituation umgehen kann. Der Beitrag endet mit einer Zusammenfassung sowie mit einer Übersicht über Forschungsbedarf in dem hier bearbeiteten Gebiet (Abschnitt 4).

## 2. Die Funktionsweise einer DEA und das Basismodell nach Charnes/Cooper/Rhodes (1978)

Wie bereits in der Einleitung erwähnt, verwendet die DEA Input-Output-Daten von Vergleichseinheiten, um eine Technologiemenge zu konstruieren. Diese Technologiemenge besitzt einen äußeren, effizienten Rand, die aus Pareto-Koopmans-effizienten Produktionen besteht. Zur Quantifizierung der Effizienz einer DMU wird der Abstand der zugehörigen Produktion zum effizienten Rand durch eine Distanzfunktion gemessen.<sup>6</sup> Ohne hier näher auf produktions- und entscheidungstheoretische Bezüge der DEA einzugehen,<sup>7</sup> sei zunächst darauf hingewiesen, dass sich die Technologiemenge formal definieren lässt durch

$$TM = \left\{ z \in R_+^{r+s} \mid z \text{ ist realisierbar} \right\}, \quad (1)$$

wobei  $r$  die Zahl der Inputs,  $s$  die Zahl der Outputs und  $z$  einen reellwertigen Vektor der Dimension  $r+s$  bezeichnet. Grundsätzlich bilden alle beobachteten DMU-Produktionen  $z_k := (x_k, y_k)_{k=1, \dots, n}$  Elemente von  $TM$ . Hierbei beschreiben  $x_k := (x_{k,1}, \dots, x_{k,r})$  den Input- bzw.  $y_k := (y_{k,1}, \dots, y_{k,s})$  den Outputvektor von DMU  $k$ . Ferner wird angenommen, dass sich sämtliche Elemente von  $TM$  durch dieselben – metrisch skalierten und nicht-negativen – Inputs und Outputs beschreiben lassen, und dass sie Realisierungen desselben, im Allgemeinen unbekanntem Produktionsprozesses sind.<sup>8</sup> Zudem wird ausgeschlossen, dass Outputs ohne den Einsatz von Inputs erzeugt werden können. Alle übrigen Eigenschaften lassen sich aus der folgenden Formalisierung herleiten:

$$TM_\xi = \left\{ z \in R_+^{r+s} \mid z = \sum_{k=1}^n \lambda_k \cdot z_k; \lambda \in \Lambda_\xi \subseteq R_+^n \right\}. \quad (2)$$

Bestimmte Linearkombinationen der Beobachtungen stellen demnach für möglich erachtete Produk-

6 Zu einer Systematisierung verschiedener Typen von Distanzfunktionen siehe Kleine (2002), S. 180 ff. und Glaser/Kleine (2004).

7 Vgl. dazu z.B. Dyckhoff/Gilles (2004) und Kleine (2002).

8 Diese beiden Annahmen, die nicht die eigentliche Konstruktion der Technologiemenge betreffen, bezeichnen Dyckhoff/Allen (1999) und Dyckhoff/Allen (2001) als Axiome.

9 Solche Linearkombinationen werden auch als virtuelle Produktionen bezeichnet.

tionen dar.<sup>9</sup> Die konkrete Beschaffenheit der Technologiemenge (insbesondere: die Skalenertragsannahme) wird durch eine entsprechende Zusatzbedingung  $\zeta$  an die DMU-bezogenen Gewichte (Linearfaktoren)  $\lambda=(\lambda_1, \dots, \lambda_n)$  erreicht.<sup>10</sup> Die Konkretisierung der Technologiemenge legt die Referenzmenge fest und hat somit maßgeblichen Einfluss auf die ausgewiesenen (In-)Effizienzen der beobachteten DMUs.<sup>11</sup>

Das DEA-Modell nach Charnes/Cooper/Rhodes (1978) unterstellt konstante Skalenerträge ( $\zeta = \text{constant returns to scale} = \text{CRS}$ ), d.h.  $\lambda_{\zeta} \in R^n$ . Grundsätzlich kann das Modell sowohl input- als auch outputorientiert formuliert werden, d.h. es ist möglich, Ineffizienzen entweder auf Input- oder auf Outputseite zu identifizieren.<sup>12</sup> Im Falle der Inputorientierung ermittelt die DEA die Effizienz einer bestimmten DMU »o« durch Lösung des folgenden Linearen Programms:

$$\begin{aligned} \min \quad & \theta_o \\ \text{unter} \quad & \theta_o \cdot x_{o,i} \geq \sum_{k=1}^n \lambda_{o,k} \cdot x_{k,i} \quad \forall i = 1, \dots, r \\ & y_{o,j} \leq \sum_{k=1}^n \lambda_{o,k} \cdot y_{k,j} \quad \forall j = 1, \dots, s \\ & \lambda_{o,k} \geq 0, k = 1, \dots, n. \end{aligned} \tag{3}$$

Das angegebene Lineare Programm (LP) sucht folglich nach einer Linearkombination aus Vergleichseinheiten, die mindestens soviel Output erzeugt wie DMU<sub>o</sub>, dafür aber nur einen Bruchteil (das  $\theta_o$ -fache) des Inputs benötigt. Wegen des unmittelbaren Bezuges zum Begriff der Technologiemenge, die sämtliche Beobachtungen umfasst (»envelopment« = Umhüllung), wird diese Modellform auch Envelopment-Form genannt. Da das Effizienzmaß  $\theta_o$  radial ist (es bezieht sich in derselben Höhe auf

sämtliche Inputs) und der Referenzpunkt  $(X\lambda_o, Y\lambda_o)$  mit den Input- bzw. Outputmatrizen  $X=(x_1, \dots, x_n)$  und  $Y=(y_1, \dots, y_n)$  nicht notwendig Pareto-Koopmans-effizient ist, wird das LP meist um einen weiteren Schritt ergänzt, bei dem nicht-radiale Ineffizienzen in Form von sog. Slacks identifiziert werden können.<sup>13</sup> DMU<sub>o</sub> ist Pareto-Koopmans-effizient genau dann, wenn  $\theta_o=1$  und sämtliche Slackwerte gleich Null sind.

Gemäß der Dualitätstheorie gibt es zu jedem LP ein assoziiertes, sog. duales LP. Variablen des einen LPs entsprechen Nebenbedingungen des anderen und umgekehrt. Ist ein LP lösbar, so auch sein Dual; die Zielwerte (hier: Effizienzwerte) stimmen überein.<sup>14</sup> Oft verwendet man in einer DEA beide Modellformulierungen, um die Ergebnisinterpretation zu erweitern. Zudem sind in jeder gängigen DEA-Software die Ergebnisse beider LPs verfügbar. Aus diesen Gründen sei auch das zum o.g. LP duale Programm angeführt:

$$\begin{aligned} \max \quad & \omega_o = \sum_{j=1}^s \mu_{o,j} \cdot y_{o,j} \\ \text{unter} \quad & \sum_{i=1}^r v_{o,i} \cdot x_{o,i} = 1 \\ & \sum_{j=1}^s \mu_{o,j} \cdot y_{k,j} \leq \sum_{i=1}^r v_{o,i} \cdot x_{k,i} \quad \forall k = 1, \dots, n \\ & \mu_{o,j} \geq 0 \quad \forall j = 1, \dots, s \\ & v_{o,i} \geq 0 \quad \forall i = 1, \dots, r. \end{aligned} \tag{4}$$

Inhaltlich besagt dieses LP Folgendes: Jede DMU »darf« die für sie optimale Gewichtung der Input- und Outputfaktoren (Effizienzkriterien) wählen; setzt man diese Gewichtung jedoch zur Bewertung aller übrigen DMUs an, soll ein Effizienzwert von maximal Eins (oder 100%) entstehen (Normierungsbedingungen). Aus offensichtlichen Gründen wird diese Modellformulierung Multiplier-Form genannt. Zudem erkennt man an dieser LP-Form, dass innerhalb der DEA die Produktivität – der Quotient aus Output und Input – als Effizienzkriterium herangezogen wird: Teilt man die linke Seite der zweiten Nebenbedingung durch die rechte Seite, erhält man den Quotienten einer gewichteten Summe von Outputs und einer gewichteten Summe von Inputs. Entscheidungstheoretisch gesehen entspricht die Gewichtung der verschiedenen Effizienzkriterien einer sog. Kompromisslösung vektorieller Entscheidungsmodelle.<sup>15</sup>

An beiden erwähnten Modellformulierungen

10 Eine Auflistung verschiedener Arten von Technologiemengen findet sich z.B. bei Kleine (2002), S. 131 bzw. Kleine (2002), S. 156 ff.  
 11 Welche inhaltlichen Überlegungen zur Konkretisierung einer Technologiemenge führen, beschreibt bspw. Allen (2002).  
 12 Neben orientierten DEA-Modellen gibt es auch sog. unorientierte Modelle, wobei das Additive Modell das gebräuchlichste dieser Modelle darstellt. Zum Additiven Modell vgl. Cooper/Seiford/Tone (2006), S. 91 ff.  
 13 Modelle mit Slackwerten finden sich in allen gängigen Lehrbüchern zur DEA; vgl. z.B. Cooper/Seiford/Tone (2006), S. 96 ff.  
 14 Da die Zielwerte von LP (3) und LP (4) nur im Optimum übereinstimmen, werden in beiden LPs unterschiedliche Variablen für den Effizienzwert verwendet.  
 15 Vgl. Kleine (2002), S. 43 ff.

lassen sich noch einmal die in der Einleitung angeführten Dimensionen der Datenqualität verdeutlichen: In der Envelopment-Form wirkt sich die Festlegung der Vergleichseinheiten auf die Zahl der Parameter – der  $\lambda$ -Werte – aus, während die Effizienzkriterien die Anzahl der Nebenbedingungen vorgeben. In der Multiplier-Form verhält es sich genau umgekehrt: Während die Effizienzkriterien die Zahl der Parameter vorgeben, legt man mit der Wahl der DMUs zugleich die Anzahl der Nebenbedingungen fest.

Auch wenn wir hier nur das grundlegende inputorientierte DEA-Modell vorgestellt haben, sei darauf hingewiesen, dass sich auf analoge Weise outputorientierte Modelle aufstellen und interpretieren lassen. Outputs stellen jedoch – im Gegensatz zu Inputs – zu maximierende Kriterien dar. Aus diesem Grund ist in den analogen outputorientiert formulierten LPs »min« durch »max« und umgekehrt zu ersetzen. Für eine ausführlichere Darstellung sei bspw. auf das Lehrbuch von Cooper/Sieford/Tone (2006) verwiesen.

### 3. Grundlegende Probleme der Datenqualität in einer DEA

#### 3.1. Problem 1: Auswahl der Vergleichseinheiten

Die Auswahl der Vergleichseinheiten ist das erste grundlegende Problem der Datenqualität. Als Ausgangspunkt dient dabei die Tatsache, dass die DEA aufgrund der nicht-parametrischen Vorgehensweise gegenüber »Ausreißern« besonders anfällig ist.<sup>16</sup> Unter Ausreißern verstehen wir solche DMUs, die die Lage des effizienten Randes und damit auch die Effizienzwerte der übrigen DMUs »in hohem Maße« beeinflussen. Allgemeine Grenzwerte für eine solche »maßgebliche« Beeinflussung gibt es jedoch nicht; vielmehr werden in empirischen Analysen häufig verschiedene Grenzwerte angesetzt, um die Abhängigkeit der DEA-Ergebnisse von einzelnen DMUs zu testen (z. B. Banker/Chang 2006). Vorerst begnügen wir uns daher mit der hier genannten Definition von Ausreißern. Ausreißer können durch fehlerhaft gemessene Inputs und/oder Outputs sowie durch Kodierungs- bzw. Übertragungsfehler entstehen. Zudem können die Inputs und Outputs bestimmter Einheiten z. B. durch zufallsbedingte Störgrößen (Rauschen) »kontaminiert« sein. Neben solchen Messfehlern (»errors-in-

variables«) können Ausreißer auch durch Stichprobenfehler (»sampling errors«) entstehen, d. h. die betrachtete Menge von Beobachtungen repräsentiert nicht in angemessener Weise die gesamte Produktionstechnologie. Auch hierdurch können Beobachtungen fälschlicherweise als effizient eingestuft werden (Cherchye/Post 2003).

Vor allem sind Ausreißer »nach oben« von Interesse, also solche DMUs, die den effizienten Rand derart nach außen verschieben, dass für einen Teil der restlichen DMUs unrealistisch niedrige Effizienzwerte entstehen. Folglich sind die Identifikation von Ausreißern und ihre anschließende Eliminierung der Schlüssel zur (endgültigen) Festlegung der Vergleichseinheiten in einer DEA.

In Abbildung 1 wird das Problem der Veränderung und gegebenenfalls der Verfälschung der Technologiemenge durch Ausreißer illustriert.<sup>17</sup> Hier sind drei Linien eingezeichnet, die – je nachdem, ob und welche Beobachtungen man als fehlerhaften Ausreißer einstuft – den effizienten Rand einer Technologie mit konstanten Skalenerträgen repräsentieren. Entscheidend für eine Einstufung als Ausreißer ist nicht die Entfernung eines einzelnen Punktes zur Masse der Daten, sondern sein Einfluss auf die Lage (oder »Ausdehnung«) der Technologie. Der mit D bezeichnete Punkt liegt weit außerhalb der durch die graue Ellipse unterlegten Punktwolke.

Gleichwohl definiert er aber eine CRS-Technologie (CRS: constant returns-to-scale), die mit einer durchgezogenen Linie markiert ist, und die für die fünf nicht mit eigenen Buchstaben bezeichneten Punkte als Referenz angesehen werden kann. Insofern ist dieser Punkt aufgrund der Tatsache, dass seine durchschnittliche Produktivität mit derjenigen der fünf »namenlosen« Punkte nahezu identisch ist, für die Effizienzbewertung anderer Beobachtungen unkritisch.

Anders ist dies bei den Punkten B und C. Keiner der Punkte bewirkt für sich genommen eine starke Veränderung der Technologie (hier als Punkt-Strich-Linie durch C eingezeichnet). Tatsächlich weisen die Beobachtungen in etwa dieselbe durchschnittliche Produktivität auf. Allerdings unterscheidet sich die durchschnittliche Produktivität dieser Technologie deutlich von der Technologie, die von der durchgezogenen Linie repräsentiert

<sup>16</sup> Vgl. Dyson et al. (2001); Wilson (1995).

<sup>17</sup> Vgl. Löber/Staat (2006).

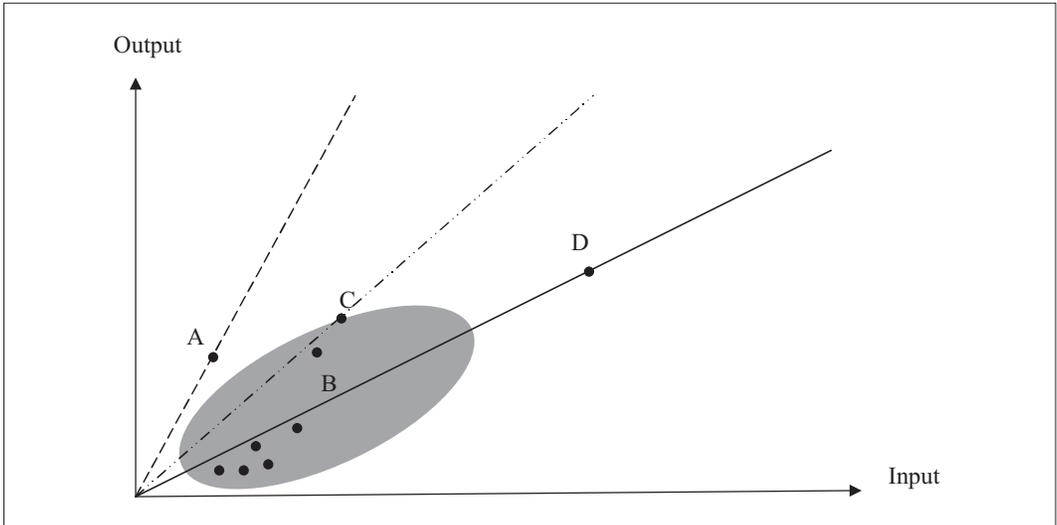


Abb. 1: Einfluss potenzieller Ausreißer auf die Technologiemenge

wird. Insofern ist zu hinterfragen, ob die beiden Punkte B und C nicht ein Paar so genannter verdeckter Ausreißer (»masked outliers«) bilden. Dies sind Beobachtungen, die allesamt Ausreißer mit einander ähnlicher Struktur sind, so dass fälschlicherweise der Eindruck entsteht, es handle sich bei ihnen um einen Teil der eigentlichen Technologiemenge.<sup>18</sup> Anders ist die Situation für den Punkt A, der den effizienten Rand der Technologiemenge deutlich nach oben verschiebt.

Die Überprüfung des Datensatzes hinsichtlich vorhandener Ausreißer ist insbesondere dann unvermeidbar, wenn es sich um große Datensätze handelt, die augenscheinlich schwer zu überblicken sind. Im Folgenden werden die wichtigsten Verfahren der Ausreißeridentifikation vorgestellt. Diese Verfahren erfordern eine Art Vorab-DEA, die DMUs aufgrund »zu hoher« Effizienzwerte aus der Effizienzanalyse ausschließt. Die eigentliche DEA wird dann ohne die identifizierten Ausreißer durchgeführt.

*Methoden der Ausreißeridentifikation*

Eine pragmatische Herangehensweise zur Ausreißeridentifikation besteht darin, zunächst eine DEA

mit dem kompletten Datensatz durchzuführen. In einem zweiten Schritt werden als effizient ausgewiesene Einheiten entfernt, und es wird auf Basis dieses reduzierten Datensatzes erneut eine DEA berechnet. Dieses Vorgehen wird so lange wiederholt, bis sich die Effizienzwerte der verbliebenen Einheiten stabilisieren.<sup>19</sup> Allerdings gibt es keine genauen Regeln, nach denen der Ausschluss effizienter Einheiten aus dem Datensatz erfolgen sollte, wodurch die Identifikation von Ausreißern zu einem nicht unwesentlichen Grad willkürlich erfolgt.

Doyle/Green (1994) bestimmen Ausreißer anhand sog. Kreuzeffizienzen.<sup>20</sup> Die Kreuzeffizienzen einer DMU *k* erhält man, indem man die Ausprägungen ihrer Inputs und Outputs mit sämtlichen optimalen DEA-Gewichten der restlichen Datenpunkte (als Lösungen der entsprechenden LPs in Multiplier-Form) bewertet. Die durchschnittliche Kreuzeffizienz (d. h. der Durchschnitt über alle einzelnen Kreuzeffizienzen) kann somit als durchschnittliche Effizienz von DMU *k* aus der Sicht aller anderen DMUs interpretiert werden.<sup>21</sup> Die Autoren formulieren aus der durchschnittlichen Kreuzeffizienz den sog. Maverick-Index

$$M_k = \frac{\theta_k - e_k}{e_k} \tag{5}$$

Dabei bezeichnet  $\theta_k$  den (wahren) Effizienzwert der Einheit *k* und  $e_k$  deren durchschnittliche Kreuzeffizienz.

18 Vgl. Löber/Staat (2006).

19 Vgl. Wilson (1995), S. 29.

20 Vgl. hierzu Doyle/Green (1994), S. 567 ff.

21 Vgl. Doyle/Green (1994), S. 569; Hammerschmidt (2006), S. 180.

zienz. Der Maverick-Index gibt die relative Differenz des DEA-Wertes zur »Fremdbewertung« auf Basis der übrigen DMUs an. Ab einem Indexwert größer 1 ist von einem Ausreißer auszugehen.<sup>22</sup> Die Bestimmung von Kreuzeffizienzen zielt primär darauf ab, Ausreißer zu identifizieren, die aufgrund extremer Konstellationen der Inputs und Outputs in der DEA extreme Gewichtungen erhalten. Der Nachteil dieser Methode besteht darin, dass Ausreißer bei vielen Datenkonstellationen nicht als solche erkannt werden.

Als weitere Methode der Ausreißeridentifikation werden in der Literatur neben den Kreuzeffizienzen auch sog. Supereffizienzen diskutiert.<sup>23</sup> Die Supereffizienz einer DMU  $k$  wird berechnet, indem sie von der Konstruktion der Technologiemenge ausgeschlossen wird. Dadurch ist ihre Effizienz  $\theta_k$  nicht mehr auf 1 (100%) normiert – in der Multiplier-Form des Linearen Programms fehlt nämlich dann die zugehörige Nebenbedingung. Während bei ineffizienten Einheiten Supereffizienz und Effizienz offenbar übereinstimmen, ist dies bei effizienten Einheiten i.d.R. nicht der Fall. Je größer die Supereffizienz einer DMU  $k$ , desto maßgeblicher ist ihr Einfluss auf die Lage des effizienten Randes und desto eher ist sie als Ausreißer zu klassifizieren. In einer Simulationsstudie zeigen Banker/Chang (2006) unter Rückgriff der von Banker/Gifford (1988) und Banker/Das/Datar (1989) entwickelten Variante der Supereffizienzberechnung, dass Supereffizienzen exaktere Effizienzschätzungen als ein herkömmliches Modell mit variablen Skalenerträgen (das sog. BCC-Modell) hervorbringen, sofern die Daten von zufälligen Messfehlern betroffen sind. Hierzu werden verschiedene Schwellenwerte von Supereffizienzen benutzt, ab denen DMUs als Ausreißer gelten (100%, 120%, 160%, 200%); außerdem wird systematisch die Wahrscheinlichkeit variiert, dass eine DMU  $k$  tatsächlich mit Messfehlern behaftet ist. Es zeigt sich, dass ein Schwellenwert von 100% – also der Ausschluss aller effizienten DMUs – zu restriktiv ist (d. h. es werden zu viele DMUs ausgeschlossen, deren Daten fehlerfrei sind), während relativ »großzügigen« Schwellen (ab 160%) zu viele messfehlerbehaftete DMUs im Datensatz verbleiben. Trotzdem lassen diese Ergebnisse keinen Schluss zu, welcher Schwellenwert grundsätzlich geeignet ist. Es gibt folglich kein formales Kriterium, das eine Entscheidung darüber zuließe, ob eine Beobachtung »signifikant auffällig« ist. Letztlich wird der kritische

Schwellenwert für den Grad der Supereffizienz, ab dem Beobachtungen aus der nachgelagerten Effizienzanalyse entfernt werden sollten, ad hoc gewählt und ggf. mit fallspezifischen Sensitivitätsanalysen kombiniert. Zudem müsste man, um auch *masked outliers* zu identifizieren, den Ausschluss der extremen Datenpunkte mehrfach durchführen. Es müssten also in einem ersten Schritt alle supereffizienten Beobachtungen, die über dem Schwellenwert liegen, aus der Referenztechnologie entfernt werden (Thanassoulis 1999). Im Rahmen einer erneuten Analyse mit dem reduzierten Referenzdatensatz müssten dann die Supereffizienz erneut berechnet und weitere Ausreißer entfernt werden. Ein solches »Abschälen« effizienter Beobachtungen (sog. peeling) reduziert möglicherweise die Referenzmenge zu stark.

Eine zuverlässigere Methode der Ausreißereliminierung stellt die Methode der sog. Order- $m$ -Schätzer dar.<sup>24</sup> Mittels der Order- $m$ -Schätzer wird

22 Vgl. Doyle/Green (1994), S. 575.

23 Als weitere Methoden zur Identifikation bzw. Berücksichtigung von fehlerbehafteten Datenpunkten werden in der Literatur gelegentlich auch die Quantilsregression (Hendricks/Koenker 1992) und das sog. »chance constrained programming« (Cooper/Huang/Li 1996) angeführt. Die Quantilsregression erlaubt über das Einführen eines Fehlerterms das Berücksichtigen zufälliger Abweichungen (bspw. in Form von Datenfehlern) beim Bestimmen der Effizienzwerte, erfordert es aber, für den Fehlerterm bestimmte Verteilungsannahmen zu treffen. Auch das chance constrained programming stützt sich auf Verteilungsannahmen, was im Zusammenhang mit einer nicht-parametrischen Methode wie der DEA fast schon als Widerspruch in sich anzusehen ist (Cherchye/Post 2003). Das Ziel dieser Methode ist es, einen »Radius der Stabilität« durch Perturbation der Originaldaten zu ermitteln, innerhalb dessen eine Variation der Input- und Outputdaten zu keiner nennenswerten Veränderung des Effizienzwertes führt. Abhängig von diesem Radius wird eine DMU dann als »wahrscheinlich effizient« oder »wahrscheinlich ineffizient« eingestuft. Insofern wäre es durch die Methode im Prinzip möglich, zulässige Grenzen von z. B. fehlerbedingten Datenvariationen zu ermitteln, die dann genutzt werden könnten, um bestimmte Punkte mit einer geringen Stabilität aus der Analyse auszuschließen. Ob ein untersuchter Punkt jedoch als mehr oder minder stabil einzustufen ist, hängt von der A-priori-Spezifikation der Varianz-Kovarianz-Matrix der Fehlerterme ab, wodurch eine starke Abhängigkeit der Resultate von den zugrunde gelegten Verteilungsannahmen erzeugt wird. Es ist jedoch schwierig, spezifische A-priori-Verteilungsannahmen gegenüber alternativen Annahmen plausibel zu rechtfertigen. Um der nicht-parametrischen Vorgehensweise der DEA Rechnung zu tragen, legen wir in diesem Beitrag Wert darauf, eine vollständig nicht-parametrische Berücksichtigung von Datenproblemen sicherzustellen. Aus diesem Grunde sehen wir von einer ausführlicheren Behandlung dieser beiden methodischen Ansätze ab.

24 Vgl. Cazals/Florens/Simar (2002), S. 2 ff.; Simar (2003), S. 391 ff.

der effiziente Rand einer Technologiemenge derart bestimmt, dass er hinsichtlich statistischer Schwankungen und Extremwerte robust ist. Der Ansatz erlaubt explizit, dass vereinzelt Datenpunkte – zu denen auch die Ausreißer gehören – oberhalb des effizienten Randes liegen. Der Abstand einer DMU zum effizienten Rand – und damit ihre (In-)Effizienz – ist nun keine deterministische Größe mehr, sondern wird auf Basis einer erwarteten Effizienzgrenze ermittelt.<sup>25</sup> Effizienzmessungen mittels der Order- $m$ -Methode erfolgen jeweils input- oder outputorientiert.

In der inputorientierten (bzw. outputorientierten) Sichtweise wird eine bestimmte DMU<sub>*o*</sub> nur mit denjenigen Einheiten verglichen, die sie outputseitig (bzw. inputseitig) dominieren, die also von jedem Output mindestens genauso viel produzieren (bzw. von jedem Input höchstens genauso viel verbrauchen). Die individuelle Referenztechnologie für eine DMU ist also eine Teilmenge der ursprünglichen Technologiemenge.

Nachdem auf diese Weise für jede zu analysierende DMU eine individuelle Referenztechnologie ermittelt worden ist, werden aus den dominierenden Technologiemengen Unterstichproben des Umfangs  $m$  gezogen.<sup>26</sup> Bei Inputorientierung bildet diejenige Einheit die erwartete untere Grenze  $m$ -ten Grades  $\theta_m$ , die innerhalb der Unterstichprobe den höchsten Effizienzwert erzeugt, d. h. das minimal erreichbare Inputniveau aufweist. Also gilt für eine bestimmte DMU<sub>*o*</sub>:<sup>27</sup>

$$\theta_m^o = E[\min(X_1, X_2, \dots, X_m) | Y \geq y_o], \quad (6)$$

wobei  $E$  den Erwartungswert (der bedingten Zufallsgröße  $\min(X_1, \dots, X_m | Y \geq y_o)$ ),  $X_j$  bis  $X_m$  den Input der Beobachtungen der Unterstichprobe,  $Y$  den Output der Einheiten im Referenzset und  $y_o$  den Output der zu bewertenden DMU<sub>*o*</sub> repräsentieren.<sup>28</sup>

In analoger Weise wird im outputorientierten Fall verfahren; hier ergibt sich:

$$\psi_m^o = E[\max(Y_1, Y_2, \dots, Y_m) | X \leq x_o]. \quad (7)$$

Ausgehend von diesen  $m$  Werten  $\psi_m$  bestimmt man nun den Effizienzwert für DMU<sub>*o*</sub>, indem in die Effizienzwerte einer Einheit nur die  $m$  gezogenen Werte aus dem dominierenden Subset eingehen. Bei Inputorientierung bedeutet dies konkret, dass die DMU<sub>*o*</sub> mit derjenigen Einheit aus der Stichprobe verglichen wird, die im Vergleich am wenigsten Input verbraucht und somit dem effizienten Rand am nächsten liegt. Analoges gilt für die Outputorientierung. Um einen stabilen Erwartungswert der Effizienzen zu erhalten, wiederholt man die Stichprobenbildung aus dem Referenzdatensatz i. d. R. mindestens  $B=200$  Mal.<sup>29</sup> Die Effizienzwerte werden als Durchschnitt aus den  $B$  Vorgängen errechnet. Beim inputorientierten Effizienzmaß erhalten ineffiziente Werte einen Wert kleiner 1, während beim outputorientierten Pendant Ineffizienzen mit einem Wert größer 1 ausgewiesen werden. Da die Effizienzmessung auf der Basis von Stichproben erfolgt, ist es möglich, dass z. B. im inputorientierten Fall Effizienzen  $>1$  ausgewiesen werden. Je größer  $m$  gewählt wird, desto seltener werden Übereffizienzen ausgewiesen.

Überschreitet im inputorientierten Fall die Übereffizienz einer DMU einen bestimmten vorher festgelegten Schwellenwert von  $(1 + \alpha)$ , so wird die entsprechende Einheit als Ausreißer klassifiziert.<sup>30</sup> Analoges gilt im outputorientierten Fall beim Unterschreiten von  $(1 - \alpha)$ . Um die passenden Werte für  $\alpha$  und  $m$  zu finden, werden in einem Diagramm für verschiedene Werte von  $\alpha$  die prozentualen Anteile der Ausreißer über  $m$  abgetragen. Je höhere Werte für  $m$  eingesetzt werden, desto niedriger wird der aus dem Datensatz ausgeschlossene Ausreißeranteil.<sup>31</sup> Tatsächlich würde dieser Anteil linear abnehmen, falls in einem Datensatz überhaupt keine Ausreißer vorhanden wären. Die Werte von  $m$ , bei denen die Kurve der fallenden Ausreißeranteile für ein bestimmtes  $\alpha$  einen Knick (»elbow«) aufweist, weisen demnach darauf hin, dass ab dort alle Ausreißer ausgeschlossen worden sind.

25 Vgl. Cazals/Florens/Simar (2002), S. 6 f.

26 Daher bezieht die Methode ihren Namen.

27 Die Formeln (6) und (7) beziehen sich auf den bivariaten Fall (1 Input, 1 Output). Für den multivariaten Fall mit einem  $r$ -dimensionalen Inputvektor bzw. einem  $s$ -dimensionalen Outputvektor muss bei radialem Effizienzmaß jeweils vor dem Vektor noch ein Maximumoperator bzw. Minimumoperator hinzugefügt werden, vgl. Simar (2003), S. 399f.

28 Es sei ausdrücklich darauf hingewiesen, dass sich hier hinter dem Symbol  $X$  Zufallsgrößen – und nicht mehr die tatsächlichen Inputs der DMUs – verbergen. Wir verwenden trotzdem für unterschiedliche Größen das gleiche Symbol, um die Komplexität der Notation im Rahmen zu halten.

29 Vgl. Löber/Staat (2006), S. 113.

30 Simar (2003) schlägt einen Schwellenwert von  $\Phi_m - 1,645 \cdot \sigma_i^{MC} \geq (1 + \alpha)$  vor, wobei  $\sigma_i^{MC}$  die Standardabweichung der Monte-Carlo-Simulation in Inputrichtung bezeichnet; vgl. Simar (2003), S. 404.

31 Für eine ausführliche Darstellung der Vorgehensweise sei auf Simar (2003), S. 403 f. verwiesen.

Es gibt keinen allgemeingültigen Wert für den Prozentsatz der DMUs, die generell als Ausreißer ausgeschlossen werden sollten. Als heuristische obere Grenze wird in der Literatur die Größe  $\frac{\sqrt{n}}{n}$

herangezogen, wobei  $n$  die ursprüngliche DMU-Anzahl bezeichnet.<sup>32</sup>

Da diese Vorgehensweise relativ willkürlich erscheint, schlagen wir zur Identifizierung von Ausreißern ein zweistufiges Verfahren vor: Im ersten Schritt werden alle DMUs als Ausreißer ausgewiesen und vom Datensatz ausgeschlossen, die sowohl in Inputrichtung als auch in Outputrichtung einen Order- $m$ -Effizienzwert über bzw. unter dem Schwellenwert aufweisen. Es ist jedoch wahrscheinlich, dass in der ersten Stufe nicht alle Ausreißer erkannt werden und weitere potenzielle Ausreißer (sog. *masked outliers*) im Datensatz verbleiben. Solche Ausreißer könnten jene Einheiten sein, die einen Order- $m$ -Effizienzwert von größer eins (im inputorientierten Fall) oder kleiner eins (im outputorientierten Fall) bei einer geringen Anzahl von Referenzeinheiten aufweisen. Die endgültige Entscheidung, ob es sich bei diesen Einheiten tatsächlich um *masked outliers* handelt, erfolgt individuell für jede Einheit, und zwar durch die Effizienzinformationen in der jeweils anderen Richtung.

Liegt z.B. der Schwellenwert bei  $\alpha = 0,3$  und weist eine DMU in Inputrichtung einen Effizienzwert von 1,1 bei einer Anzahl von 5 Einheiten im Referenzset auf, so wird diese Einheit im ersten Schritt nicht als Ausreißer identifiziert. Allerdings könnte es sich bei dieser Einheit dennoch um einen potenziellen Ausreißer handeln, da sie aufgrund ihrer hohen Effizienz und der geringen Anzahl von dominierenden Einheiten extreme Tendenzen aufweist. Diese Einheit wird dann im zweiten Schritt als Ausreißer entfernt, wenn sie auch in Outputrichtung einen Effizienzwert von über eins bei kleinem Referenzset aufweist. Weist jedoch die betrachtete Einheit in Outputrichtung z.B. einen Effizienzwert von 1,3 bei 300 Referenzeinheiten aus, kann davon ausgegangen werden, dass es sich nicht um einen *masked outlier* handelt.

### Empirische Illustration

Die Anwendung des Order- $m$ -Verfahrens erfolgt für Daten von 324 Händlern eines deutschen Automobilunternehmens. Es wird betrachtet, wie effizient ein Händler mit seinen Mitarbeitern (Anzahl

der Mitarbeiter im Service, Anzahl der Mitarbeiter in der Werkstatt, Anzahl der Mitarbeiter im Teileverkauf) und Kapitalressourcen (Anzahl der Reparaturplätze) aus dem gegebenen Fahrzeugbestand After-Sales-Umsätze im Reparaturgeschäft generiert. Die händlerindividuellen Fahrzeugbestände bilden das Marktpotenzial bzw. Kundenpotenzial eines Händlers ab. Die Fahrzeugbestände wirken sich über die Generierung von Kundendurchgängen pro Händler auf den Umsatz aus. Es ist davon auszugehen, dass prinzipiell nur für einen bestimmten Prozentsatz des Fahrzeugbestandes After-Sales-Transaktionen pro Jahr erzielt werden können. Daher ist die Berücksichtigung der Größe »Kundendurchgänge« als weiterer Input für eine aussagekräftige Effizienzbewertung wichtig. Als sog. nicht diskretionäre Variable wird die Wettbewerbsintensität im Einzugsgebiet eines Händlers betrachtet, da davon auszugehen ist, dass eine hohe Wettbewerberzahl – für einen gegebenen Input – in geringerer Output und somit geringerer Effizienz resultiert.<sup>33</sup>

Zunächst fällt die extrem große Streuung der Order- $m$ -Effizienzwerte auf. Es gibt also sowohl viele extrem ineffiziente Einheiten als auch viele Einheiten mit großen Übereffizienzen. Dies ist ein erster Hinweis darauf, dass die DMUs hinsichtlich ihrer Input-Output-Konstellationen relativ weit auseinander liegen, d.h. wenige Referenzeinheiten besitzen, mit denen sie verglichen werden können.

Die hohe Anzahl extrem effizienter Einheiten ist auch die Ursache dafür, dass bei der Ausreißerbestimmung vergleichsweise viele Einheiten als solche klassifiziert werden. Das wird auch in den Abbildungen 2 und 3 deutlich, in denen die Ausschusskurven des Order- $m$ -Verfahrens bei Input- und Outputorientierung jeweils unter Zugrundelegung variabler Skalenerträge (BCC-Modell) dargestellt sind. Wie bereits erläutert, stehen die einzelnen Kurven für den Prozentsatz der Einheiten, die bei einem bestimmten Wert für  $m$  bei dem jeweiligen Schwellenwert  $\alpha$  als potenzielle Ausreißer identifiziert werden. So würden bei  $\alpha = 0,2$  und  $m = 20$  bei Inputorientierung ca. 27% der Einheiten als potenzielle Ausreißer eingestuft.

Zur Festlegung von  $m$  und  $\alpha$  wird in den Abbildungen nach den im vorangegangenen Absatz er-

<sup>32</sup> Vgl. Simar (2003), S. 404.

<sup>33</sup> Im Rahmen des zweiten Problems – Wahl der Effizienzkriterien – werden wir auf die Problematik nicht diskretionärer Variablen ausführlicher eingehen.

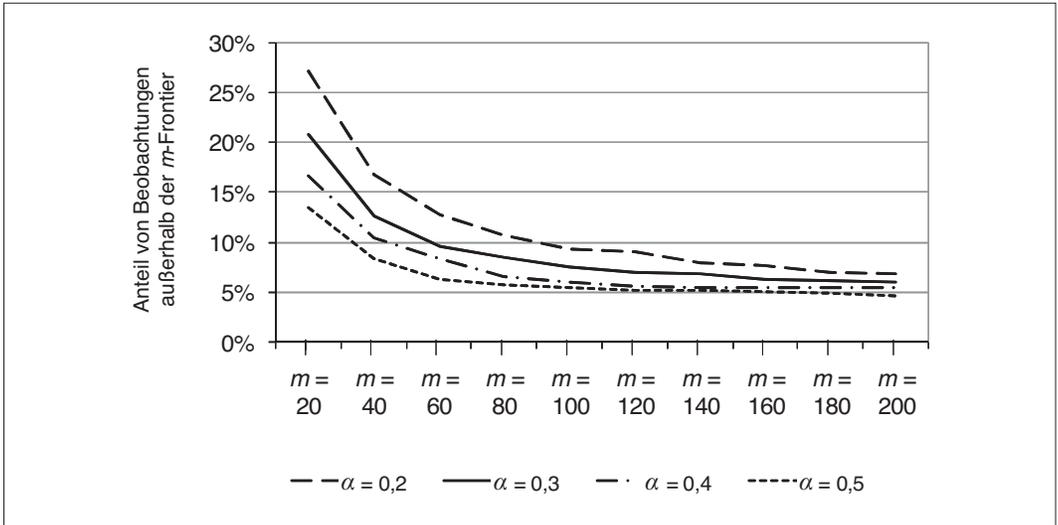


Abb. 2: Order-m-Ausschlusskurven, inputorientiert

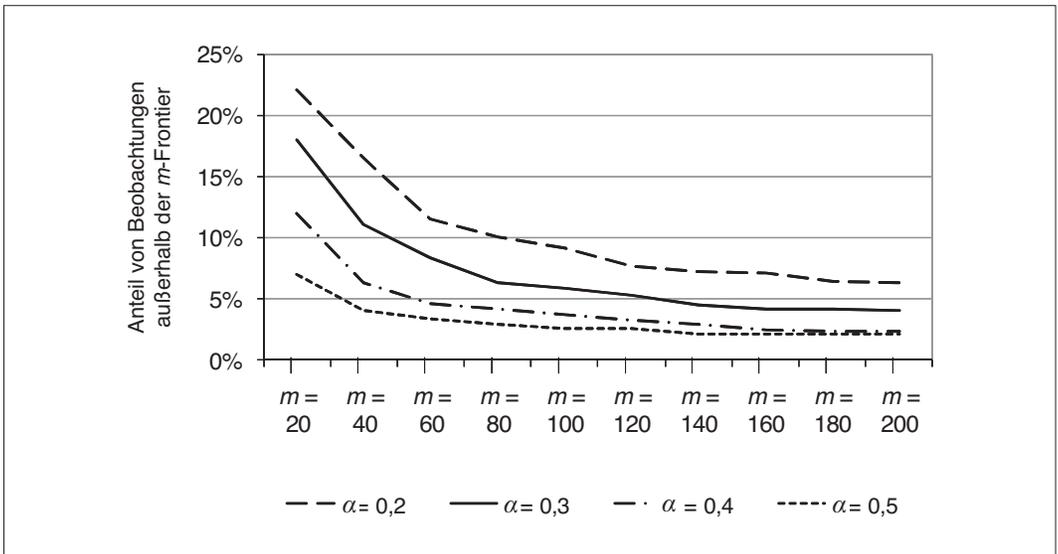


Abb. 3: Order-m-Ausschlusskurven, outputorientiert

wählten Knickstellen gesucht. Da einige Händler auch Reparatur-Services für andere Betriebe der Vertriebsorganisation durchführen und somit wahrscheinlich a priori überdurchschnittliche Umsätze aufweisen, die nicht in erster Linie auf besonders effizientes Handeln zurückzuführen sind, sollten in dieser Studie eher zu viele als zu wenige Ausreißer ausgeschlossen werden.

Zwar weisen die Ausschlusskurven des Order-m-Verfahrens mehrere Knickstellen auf, jedoch kommen diese in beiden Richtungen vermehrt bei  $m = 40$  vor. Daher wird  $m = 40$  als Umfang der Unterstichproben für die Ausreißerbestimmung herangezogen. Der Schwellenwert wird auf  $\alpha = 0,2$  festgesetzt, da der Anteil potenzieller Ausreißer der oben genannten Zielsetzung entsprechend mit

Durchschnittlicher Effizienzwert	Anzahl betrachteter Einheiten	Anzahl effizienter Einheiten	Effiziente Einheiten in Prozent
327,80 %	297	44	14,8 %

Tab. 1: Effizienzergebnisse des outputorientierten DEA-Modells nach Ausreißereliminierung

16,4% in Inputrichtung und 16,1% in Outputrichtung hoch liegt. Im ersten Schritt werden nur die DMUs ausgeschlossen, die in Input- und Outputrichtung als Ausreißer identifiziert werden. Der Ausreißer-Anteil kann daher bei höchstens 16,1% liegen. Dies wäre der Fall, wenn die Überschneidung maximal wäre. Für unsere Studie beträgt die Schnittmenge nur 6,7%, weshalb in der ersten Stufe 22 DMUs als Ausreißer eliminiert werden. Wie oben beschrieben werden nun in einer zweiten Stufe auch noch solche DMUs als Ausreißer entfernt, die in beiden Richtungen bei kleinem Referenzset<sup>34</sup> einen Effizienzwert größer (inputorientiert) bzw. kleiner (outputorientiert) eins aufweisen. In der zweiten Stufe werden so nochmals 5 DMUs (1,5%) als masked outliers eliminiert. Insgesamt werden somit 27 DMUs entfernt, was mit 8,3% dem Eineinhalbfachen des heuristischen Richtwertes von  $\frac{\sqrt{n}}{n} = \frac{\sqrt{324}}{324} \approx 5,6\%$  entspricht. Der höhere

Wert ist im Hinblick auf die erwünschte strenge Vorgehensweise gerechtfertigt.

Nach Eliminierung der 27 Ausreißer wird ein outputorientiertes DEA-Modell mit variablen Skalenerträgen ( $\zeta$  = variable returns to scale = VRS) für die verbleibenden 297 Einheiten verwendet.

Ohne Ausreißereliminierung wird für ca. 80% der Händler eine zu hohe Ineffizienz ausgewiesen. Diese Händler werden somit ohne Ausreißerbereinigung als zu schlecht und somit unfair bewertet. Dies hätte vermutlich zu starken und berechtigten Widerständen gegen die Performancebewertung mittels DEA geführt. Während im Ausgangsmodell nur 23 Händler und somit 7,1% der Händler als effizient ausgewiesen werden (d. h. einen Effizienzwert von 100% aufweisen), steigt diese Zahl im bereinigten Modell auf 44 Händler (14,8%; vgl. Tabelle 1). Die durchschnittliche Effizienz ist auch nach Ausreißereliminierung mit 327,80% sehr niedrig. Dieser Wert besagt, dass Referenzhändler (»best practices«) existieren, die mit dem gleichen Input mehr als dreimal so viel Output generieren.

Die Standardabweichung weist auf eine große Streuung der Effizienzwerte hin. Folglich muss man davon ausgehen, dass der Datensatz viele sehr stark ineffiziente Einheiten enthält. Dies ist allerdings insofern unkritisch, als die extrem ineffizienten Einheiten lediglich selbst einen niedrigen Effizienzwert aufweisen, jedoch nicht die Effizienzwerte der übrigen Händlerbetriebe beeinflussen. Allerdings muss man sich bei extrem ineffizienten Händlern darüber im Klaren sein, dass die identifizierten Verbesserungspotenziale unter Umständen nur sehr schwer ausgeschöpft werden können. In Bezug auf den mit dem Fahrzeugbestand maximal erreichbaren Reparaturumsatz zeigt sich, dass die Händler annähernd 100% Marktanteil bei Reparaturen bezogen auf die Fahrzeugmodelle des betrachteten Unternehmens aufweisen müssten, um die Ineffizienz vollständig abzubauen. Allerdings kann angesichts der Tatsache, dass einige Händler durchaus Marktanteile von nahe 100% bei Reparaturen erreichen, die Aussage getroffen werden, dass eine durchschnittliche Effizienzerhöhung pro Händler von 327% durchaus als erreichbar anzusehen ist und daher nicht außerhalb des Machbaren liegt.

### 3.2. Problem 2: Auswahl der Effizienzkriterien

Die Wahl der Effizienzkriterien als zweites grundlegendes Problem der Datenqualität in einer DEA beinhaltet zwei Aspekte: Wenn aus konzeptionellen Überlegungen heraus die zur Beschreibung des Produktionsprozesses notwendigen Inputs und Outputs feststehen, ist zu überprüfen, ob sie tatsächlich die Effizienz beeinflussen. Gleichzeitig muss geklärt werden, welche der ausgewählten Kriterien tatsächlich im Einflussbereich der DMUs stehen, was zu einer Unterteilung der Variablen in

<sup>34</sup> Die Grenze, ab der ein Referenzset als klein gilt, wurde konservativ mit 10 Einheiten festgelegt.

diskretionäre (frei variierbare, kontrollierbare, beeinflussbare) sowie nicht diskretionäre (nicht variierbare, nicht kontrollierbare, nicht beeinflussbare) Variablen führt. Nicht diskretionäre Variablen beschreiben häufig das Umfeld, in dem die DMUs operieren, weswegen sie in dem Fall auch als »environmental parameters« (Umfeldvariablen) bezeichnet werden. Sie dürfen nicht in die Zielfunktion zur Bestimmung des Effizienzmaßes einfließen, da die Verantwortlichen nicht in der Lage wären, die auf dieser Basis abgeleiteten Zielgrößen zu realisieren. In den Nebenbedingungen müssen sie jedoch als zusätzliche Restriktionen integriert werden. Dementsprechend enthält das DEA-Modell, welches zur Bestimmung der Händlereffizienz in Abschnitt 3.1 verwendet wurde, eine zusätzliche Nebenbedingung, die die Wettbewerbsintensität berücksichtigt. Es ist zu beachten, dass die Effizienzwerte gemäß der gewählten Orientierung des DEA-Modells entweder nur durch die Inputs oder nur durch die Outputs beeinflusst werden und somit nicht diskretionäre Variablen auf der jeweils anderen Seite unkritisch sind.

Je nachdem, ob man einen Input oder Output als diskretionär oder nicht diskretionär einstuft, stehen unterschiedliche Verfahren zur Verfügung, um ihre Relevanz bzw. Signifikanz im Rahmen der Effizienzmessung zu testen. Während es bei nicht diskretionären Variablen einen formal-analytischen Signifikanztest gibt, den wir ausführlich erläutern werden, bleiben für den Fall diskretionärer Variablen heuristische Verfahren, um zumindest Hinweise auf die Relevanz der potenziellen Input- und Outputvariablen zu erhalten.

#### *Methoden zur Überprüfung der Relevanz diskretionärer Variablen*

Zunächst sei darauf hingewiesen, dass der Ausschluss eines Inputs oder Outputs aufgrund seiner hohen Korrelation mit einem anderen Input oder Output problematisch ist.<sup>35</sup> Eine andere Möglichkeit zur Überprüfung diskretionärer Variablen im Hinblick auf ihre Beeinflussung der Effizienzwerte besteht darin, die Ergebnisse eines Modells mit allen potenziell relevanten Variablen (»full model«) mit den Ergebnissen von Teilmodellen, in denen je-

weils eine Variable entfernt wird (»omitted variable models«), zu vergleichen, um die Sensitivität des Modells abzuschätzen. Die jeweiligen Korrelationen zwischen den Effizienzwerten des vollständigen Modells und denen der »omitted models« zeigen dann, bei welchen Variablen das Weglassen einen erheblichen Einfluss auf die Höhe der Effizienzwerte ausübt.<sup>36</sup> Die Variablen, bei deren Weglassen die Effizienzwerte besonders sensitiv reagieren, sind als besonders kritisch und einflussreich anzusehen. Insbesondere in dem Bereich, aus dem die besonders einflussreichen Variablen stammen, ist nach weiteren potenziell relevanten Größen zu suchen. Die wichtigste Grundlage bildet hierfür das Zielsystem des Anwenders. Um einen kausalen Zusammenhang zwischen Inputs und Outputs unterstellen zu können, ist abschließend sicherzustellen, dass die Input- und Outputvariablen positiv (zumindest aber nicht negativ) korreliert sind.<sup>37</sup>

#### *Methoden zur Überprüfung der Signifikanz nicht diskretionärer Variablen*

Für das Berücksichtigen von nicht diskretionären Variablen finden sich in der Literatur verschiedene Modellierungsansätze. Allen Ansätzen ist zunächst gemein, dass durch die Formulierung entsprechender Nebenbedingungen im LP – wie oben schon am Beispiel der Wettbewerbsintensität erläutert – die Referenztechnologie auf dasselbe Niveau justiert wird, das die nicht diskretionäre Variable der zu bewertenden  $DMU_o$  aufweist. So können nur solche DMUs zum Vergleich herangezogen werden, bei denen bzgl. der nicht diskretionären Variablen dieselben Werte wie bei  $DMU_o$  vorliegen.<sup>38</sup> Für eine nicht diskretionäre Variable muss hier zum einen bekannt sein, ob eine die Effizienz steigernde Wirkung von ihr ausgeht oder das Gegenteil der Fall ist; zum anderen muss ihre generelle Relevanz zur Beschreibung des Produktionsprozesses bestimmt werden. Bei sog. einstufigen Ansätzen wird die Relevanz der fraglichen nicht diskretionären Variable(n) für die Bestimmung der Effizienzwerte lediglich angenommen, jedoch nicht vorab durch einen Signifikanztest geprüft.

Daher besteht die Notwendigkeit, irrelevante Variablen mit einem lediglich vermuteten oder behaupteten Einfluss auf die Leistung der  $DMU_o$  zu identifizieren und diese in der eigentlichen Effizienzbewertung nicht mehr zu berücksichtigen. Ziel ist es also, vor Durchführung der endgültigen DEA zu prüfen, welche der nicht diskretionären Variab-

35 Vgl. Pedraja-Chaparro/Salinas-Jimenez/Smith (1999).

36 Vgl. Donthu/Hershberger/Osmonbekov (2005), S. 1480.

37 Vgl. Dyson et al. (2001), S. 248.

38 Vgl. für die formale Darstellung der Modellansätze ausführlich Staat (1999).

len in einem signifikanten Zusammenhang mit den Effizienzwerten stehen. Genau dieses Ziel verfolgen die sog. zweistufigen Verfahren. Auf einer ersten Stufe wird ein DEA-Basismodell mit ausschließlich diskretionären Variablen berechnet. Die so abgeleiteten Effizienzwerte werden auf einer zweiten Stufe auf nicht diskretionäre Größen regressiert. Nur die nicht diskretionären Größen mit signifikantem Einfluss werden dann im endgültigen DEA-Modell über die oben beschriebene Erweiterung der Nebenbedingungen integriert.

Alle bisher publizierten Studien, die einem zweistufigen Ansatz folgen, nutzen auf der zweiten Stufe ausnahmslos *Ad-hoc-Lösungen* in Form von OLS- bzw. Tobit-Regressionen.<sup>39</sup> Der OLS-Ansatz geht fehl, da die Effizienzwerte als abhängige Variable auf der zweiten Stufe beim Wert 1 beschränkt sind (vgl. Abschnitt 2). Das Tobit-Verfahren setzt eine zensierte (»censored«) abhängige Variable voraus, während es sich bei den Effizienzwerten um eine gestutzte (»truncated«) Variable handelt.<sup>40</sup> Die *Ad-hoc-Lösungen* berücksichtigen weiter nicht, dass die Effizienzschätzung, die ja die abhängige Variable auf der zweiten Stufe darstellt, verzerrt ist, da die Effizienzwerte anhand der *Schätzung* einer nicht beobachtbaren »wahren« Frontier bestimmt werden. Ebenso wird im Rahmen dieser Ansätze ignoriert, dass die Schätzwerte für die einzelnen Beobachtungen untereinander über die gemeinsame Referenztechnologie korreliert sind, und dass folglich auch der Störterm der Regression auf der zweiten Stufe mit den nicht diskretionären Variablen korreliert ist. Simar/Wilson (2007) legen die methodischen Schwächen dieser Ansätze ausführlich dar und zeigen im Rahmen von Simulationen, dass diese zu nicht validen Ergebnissen führen.

Im Gegensatz zu den *Ad-hoc-Verfahren* handelt es sich bei dem im Folgenden vorzustellenden zweistufigen Bootstrapping-Ansatz, zu dem es derzeit keine methodische Alternative gibt, um eine von Simar/Wilson (2007) entwickelte Erweiterung des Standardansatzes der DEA. Das Verfahren ermöglicht es, auf einen Zusammenhang zwischen den nicht von den DMUs zu beeinflussenden Variablen und deren Effizienzwerten zu testen. Dabei ist es nicht notwendig, a priori festzulegen, ob es sich bei der betreffenden Größe um einen Input oder um einen Output handelt – eine nützliche Eigenschaft des Verfahrens, denn oftmals steht nicht a priori fest, ob eine nicht diskretionäre Variable die Effizienz fördert oder ihr abträglich ist. Gleich-

zeitig würde die Berücksichtigung einer Vielzahl nicht diskretionärer Größen Effizienzpotenziale mit hoher Wahrscheinlichkeit überdecken. Daher ist es notwendig, die Richtung sowie die Signifikanz der Wirkung dieser Faktoren zu bestimmen, bevor sie Eingang in das endgültige DEA-Modell finden. Das Bootstrap-Verfahren leistet beides.

Das Verfahren basiert auf der Annahme, dass sich die Verteilung der Effizienzwerte durch eine Dichtefunktion beschreiben lässt, bei der die Effizienzeffekte der Inputs und Outputs der Technologie auf der einen Seite sowie der nicht diskretionären Variablen auf der anderen Seite separiert werden können. Bezogen auf die nicht diskretionären Variablen bedeutet das, dass sie zwar den Grad der Effizienz der DMUs beeinflussen, nicht aber die Form des effizienten Randes. Derjenige Faktor, der den Effekt der nicht diskretionären Variablen auf die Effizienz beschreibt, wird im Rahmen eines regressionsanalytischen Ansatzes mit gestutzter Fehlergröße parametrisiert. Formal kann dies geschrieben werden als

$$\mathcal{J}_k = \gamma' w_k + \zeta_k, \quad (8)$$

wobei  $\mathcal{J}_k$  das Effizienzniveau der DMU  $k$ ,  $\gamma$  einen (Spalten-)Vektor von Regressionsparametern (und  $\gamma'$  die Transposition von  $\gamma$ , also den entsprechenden Zeilenvektor) und  $w_k$  die nicht diskretionären Variablen beschreibt.  $\zeta_k$  ist eine i.i.d. verteilte Zufallsvariable, die unabhängig von  $w_k$  ist; die Verteilung ist von links bei  $1 - \gamma' w_k$  für alle  $k$  gestutzt.

Als Regressoren dienen somit die nicht diskretionären Variablen, die abhängige Variable ist die Schätzung der Effizienzwerte aus einem Basismodell ohne Einbeziehung nicht diskretionärer Variablen, das – wie oben beschrieben – die erste Stufe darstellt. Für die Bestimmung der Regressionsparameter auf der zweiten Stufe (d.h. zur Schätzung von  $\gamma$ , die üblicherweise mit  $\hat{\gamma}$  bezeichnet wird) kommt ein Bootstrapping-Algorithmus zum Einsatz.<sup>41</sup> Hierfür sind im Wesentlichen zwei Algorithmus-Schleifen zur Erzeugung von Bootstrap-Parameterschätzungen erforderlich, die  $B_1$  bzw.  $B_2$ -mal

39 Vgl. etwa die Studien von Isik/Hassan (2002) und Mukherjee/Ray/Miller (2001).

40 Vgl. zum Unterschied zwischen zensierten und trunkierten Variablen Simar/Wilson (2007).

41 Vgl. Simar/Wilson (2007), insbesondere die Ausführungen zu Algorithmus #2.

durchlaufen werden. Für  $B_1$  wird von Simar/Wilson (2007) 200 vorgeschlagen, für  $B_2$  1000. Durch das Erzeugen von Bootstrap-Parameterschätzungen aus Unterstichproben wird es möglich, eine (neue) Verteilung der geschätzten Parameter  $\hat{\gamma}$  der gestutzten Regression zu erzeugen. Diese Verteilung wird dazu genutzt, ein Konfidenzintervall zu bestimmen. Zunächst werden die ermittelten Parameterschätzungen in aufsteigender Reihenfolge sortiert; anschließend lassen sich aus der resultierenden Verteilung die interessierenden Perzentile bestimmen. Bei einem zweiseitigen Test auf 5%-Niveau werden die 2,5% niedrigsten bzw. höchsten Werte entfernt. Sollte sich unter den verbleibenden 95% der Wert Null befinden, lässt sich die Ausgangshypothese, dass der betreffende Faktor keinen signifikanten Einfluss auf die Effizienz ausübt, nicht verwerfen. Der skizzierte Algorithmus bedingt einen hohen Implementierungsaufwand, stellt aber gleichzeitig das einzige konsistente zweistufige DEA-Verfahren dar.

#### *Empirische Illustration*

Wir zeigen die Bedeutung des Signifikanztests im Rahmen einer DEA anhand von Daten einer deutschen Großbank. Neben den diskretionären Inputs Personal-, Sachkosten sowie Kosten für die Räumlichkeiten und den diskretionären Outputs (den in den Kern-Geschäftsfeldern Kredit, Einlagen, Wertpapiergeschäft sowie Verkauf von Versicherungen erzielten Erträgen) bezieht die Bank auch fünf nicht diskretionäre Variablen in die Leistungsbewertung ein: Kundenpotenzial, Kaufkraft, Zahl der Geldautomaten, Zahl der Geldautomaten der Konkurrenz, Öffnungsdauer. Die Bank geht von der Relevanz dieser Größen für die Effizienz der Filialen aus und bezieht diese daher im Rahmen der Vertriebssteuerung in das Bewertungs- und Vergütungssystem ein, ohne aber jemals Stärke und Richtung des Einflusses der Variablen auf die Effizienz getestet zu haben. Werden aber eine oder mehrere für die Effizienzbewertung irrelevante Variablen einbezogen, wird die Effizienz der Filialen unter Umständen überschätzt. Mögliche Verbesserungspotenziale könnten somit verdeckt werden. Dies wird anhand der folgenden Ergebnisse gezeigt. Wir ermitteln zunächst die Effizienzwerte der 119 Filialen für ein Modell, das sämtliche verfügbaren diskretionären und nicht diskretionären Variablen umfasst. Schließlich verwenden wir ein Modell, das neben den diskretionären Variablen nur jene nicht diskre-

tionären Größen beinhaltet, deren Einfluss nach den Ergebnissen der Bootstrapping-Regression auf der zweiten Stufe signifikant ist.

Werden sämtliche nicht diskretionären Variablen als Inputs behandelt und in das Modell einbezogen (Modell I), ergeben sich 73 effiziente Filialen. Der Anteil ineffizienter Filialen beträgt lediglich 38%. Letztere weisen eine durchschnittliche Effizienz von 110,84% auf, die über alle Filialen gemittelte Effizienz liegt bei 106,3%. Es verbleiben keine Filialen mit einem Effizienzwert über 150%. Das anhand von Modell I für das Filialnetz ausgewiesene Ertragspotenzial – berechnet als die Summe der Verbesserungsmöglichkeiten bzgl. sämtlicher Outputs über alle DMUs – beträgt 15,3 Mio. €.

Die gestutzte Regression auf der zweiten Stufe (vgl. Formel (8) oben) ergibt, dass lediglich die Variable Kundenpotenzial einen signifikanten Effekt aufweist; hier ist der Wert der oberen sowie der unteren Schranke des 95%-Konfidenzintervalls negativ. Bei allen anderen Parametern ist der 2,5%-Wert der Parameter negativ, während der 97,5%-Wert positiv ist, so dass das Intervall den Wert Null beinhaltet.

Folglich sollte lediglich die Variable Kundenpotenzial als einzige nicht diskretionäre Größe im DEA-Modell berücksichtigt werden (Modell II). Alle übrigen nicht diskretionären Variablen sollten aufgrund ihres nicht-signifikanten Einflusses auf die Effizienzwerte aus dem Modell ausgeschlossen werden.

Die Ergebnisse von Modell II weichen von Modell I erheblich ab: Statt 73 sind nun nur noch 53 Filialen effizient, und der durchschnittliche Effizienzwert beträgt statt 106,3% nun 111,8%. Vergleicht man die Ergebnisse der validen Spezifikation II mit Modell I, so ergibt sich, dass für 20 Filialen Effizienzdefizite ausgewiesen werden, für die anhand von Modell I kein Handlungsbedarf erkannt wurde. Teilweise werden Beobachtungen in Modell I als effizient ausgewiesen, die bei der anderen Spezifikation ein erhebliches Optimierungspotenzial – bis über 17% – aufweisen. Diese Optimierungspotenziale wären nicht aufgedeckt worden, wenn die Zahl der berücksichtigten nicht diskretionären Variablen nicht durch das zweistufige Verfahren sinnvoll beschränkt worden wäre. Insgesamt wird mit der korrekten Modellspezifikation ein fast doppelt so hohes Ertragssteigerungspotenzial identifiziert (27,1 Mio. € im Vergleich zu

15,3 Mio. €). Man sieht somit, dass die Auswirkungen einer Fehlspezifikation auf die Ergebnisse aufgrund eines fehlenden Signifikanztests beachtlich sein können.

### 3.3. Problem 3: Verhältnis der Zahl der Effizienzkriterien zur Zahl der Vergleichseinheiten

Zusätzlich zu den bisher angeführten Problemkreisen der Datenqualität bzgl. Effizienzkriterien (Inputs und Outputs) *oder* Vergleichseinheiten ist ein weiteres zentrales Problemfeld zu nennen, das sich auf die Dimension der Ursprungsdatenmatrix *insgesamt* bezieht, also auf das Verhältnis der Anzahl der Vergleichseinheiten und der Anzahl verwendeter Effizienzkriterien.

Es ist offensichtlich, dass die Zahl der Kriterien, die in einem DEA-Modell sinnvollerweise verwendet werden können, durch die Zahl der zur Verfügung stehenden Beobachtungen nach oben begrenzt ist. Anhand der Envelopment-Form des Basismodells (vgl. Abschnitt 2) ist leicht zu erkennen, dass mit der Zahl der Inputs und Outputs die Zahl der Nebenbedingungen im Optimierungsproblem steigt; schließlich muss für jede weitere Variable die Bedingung, dass die Referenztechnologie die zu bewertende Beobachtung (schwach) dominiert, zusätzlich erfüllt sein. Folglich wird  $\theta$  im inputorientierten (outputorientierten) Modell steigen (fallen). Im Extremfall sind sämtliche Vergleichseinheiten effizient, was eine effizienzorientierte Diskriminierung unmöglich macht. Im Sinne der Multiplier-Form bedeutet dies, dass aufgrund der Vielzahl an Effizienzkriterien jede Vergleichseinheit das beste Verhältnis irgendeines Outputs und irgendeines Inputs aufweist.

Umgekehrt – also bei gegebener Zahl an Effizienzkriterien – weisen Zhang/Bartels (1998) anhand einer Vergleichsstudie dreier Datensätze von Stromversorgern darauf hin, dass die durchschnittliche Effizienz umso niedriger ist, je mehr DMUs man betrachtet. Sie schlagen eine Monte-Carlo-Simulation vor, um den Umfang großer Stichproben auf die Zahl der Beobachtungen des kleinsten Datensatzes zu reduzieren und so durchschnittliche Effizienzwerte zu erhalten, die über die verschiedenen Stichproben vergleichbar sind. Die drei von ihnen untersuchten Gruppen von Stromversorgern weisen ähnliche Effizienzniveaus auf, sobald alle Datensätze auf die Zahl der Beobachtungen des

kleinsten Datensatzes beschränkt werden.

Pedraja-Chaparro/Salinas-Jimenez/Smith (1999) illustrieren dieselben Zusammenhänge anhand von Simulationen und zeigen, dass dasselbe durchschnittliche Effizienzniveau wie im Ausgangsmodell nach Hinzunahme eines weiteren Parameters im Mittel erst bei einer doppelt so hohen Anzahl von Beobachtungen wieder erreicht wird. Dies bedeutet, dass die Zahl der erforderlichen Beobachtungen mit der Zahl der Kriterien exponentiell ansteigt.

Diese zunächst als relativ praxisfern anmutenden Überlegungen weisen eine für die Interpretation der DEA-Resultate nicht zu unterschätzende Relevanz auf. So bezeichnen beispielsweise Paradi/Vela/Yang (2004) die in ihrer Studie untersuchten Bankfilialen als »well managed« und beziehen sich bei dieser Aussage auf den Umstand, dass bei entsprechenden Untersuchungen häufig der überwiegende Anteil der Filialen als effizient bestimmt wird. Dass viele Studien über Banken eine hohe Zahl effizienter Beobachtungen aufweisen, liegt aber zu einem erheblichen Teil daran, dass in diesen Studien in der Regel ein üppiger Datenkranz einer relativ geringen Anzahl von Beobachtungen (Filialen) gegenüber steht.<sup>42</sup> Dies wird insbesondere dann der Fall sein, wenn die Filialverantwortlichen daran interessiert sind, eine große Anzahl von Kriterien, die aus ihrer Sicht für die Effizienz »ihrer« Filialen von Belang sind, im Modell berücksichtigt zu wissen.

Paradi/Vela/Yang (2004) empfehlen folgende Vorgehensweise: Bei der Untersuchung eines Filialnetzwerkes sollten mehrere Modelle parallel zum Einsatz kommen. Zunächst ist ein Modell zu rechnen, das die Effizienz der betroffenen Filialen aus Sicht der Filialverantwortlichen positiv (effizient) darstellt, um Akzeptanz für die Methode zu erzeugen. Dieses kann alle potenziell relevanten Input- und Outputgrößen enthalten. Ein weiteres Modell sollte so ausgelegt sein, dass noch gut erreichbare Zielvorgaben ermittelt werden, und ergibt dann eine Art mittlere Effizienz. Schließlich wird ein drittes Modell auf Basis vergleichsweise weniger Kriterien berechnet, das das maximale Verbesserungspotenzial aufzeigt. Dieses Vorgehen lässt allerdings die Frage nach der wahren Effizienz der

<sup>42</sup> Vgl. den Überblick über DEA-Studien im Bankbereich bei Paradi/Vela/Yang (2004).

Filialen unbeantwortet, ebenso wie die Frage, welche Kriterien die Technologie konstituieren.

#### Methoden für den Test und die Korrektur von Stichprobenbedingten Verzerrungen

Mit den obigen Ausführungen wurde gezeigt, dass auf der DEA basierende Effizienzschatzungen zwangsläufig verzerrt sind. Die Schätzung der Technologiemenge  $\hat{TM}$  ist eine *innere* Approximation von  $TM$ , woraus automatisch eine Verzerrung der geschätzten Effizienzwerte resultiert (vgl. Formel (3) in Abschnitt 2). Mit einer endlichen Zahl an Beobachtungen werden die wahren technologischen Möglichkeiten unterschätzt. Somit wird es Punkte geben, an denen Produktion theoretisch möglich wäre, es aber keine vergleichbare DMU innerhalb der Daten gibt, so dass dieser Punkt »automatisch« effizient (»efficient by default«) ist.<sup>43</sup> Dieser Umstand hat zu verschiedenen Faustregeln bezüglich der Beziehung zwischen der Anzahl der Beobachtungen und der Anzahl der Effizienzkriterien geführt. So schlagen bspw. Cooper/Seiford/Tone (2006) vor, dass die Stichprobe mindestens dreimal so viele Beobachtungen wie verwendete Kriterien (Inputs und Outputs) umfassen sollte. Dyson et al. (2001) fordern, dass die Anzahl der Beobachtungen mindestens zweimal das Produkt aus der Anzahl der Inputs und der Anzahl der Outputs betragen sollte.

Simar/Wilson (2000) geben unter der Annahme einer monotonen, konkaven Produktionsfunktion eine Formel für die asymptotische Verzerrung an, die von der Stichprobengröße  $n$  sowie von »der Krümmung des effizienten Randes und der Dichte der Beobachtungen am Rand«<sup>44</sup> abhängt. Es ist intuitiv klar, dass die Verzerrung mit der Dichte ab- und mit der Krümmung zunimmt. In großen Stichproben mit einer hohen Dichte an Beobachtungen am effizienten Rand und bei gleichzeitig schwacher Krümmung ist daher eine relativ geringe Verzerrung zu erwarten. Ist die Stichprobe hingegen klein, weist der effiziente Rand Knickstellen auf und ist die Dichte von Beobachtungen am effizienten Rand gering, so ist mit einer merklichen Verzerrung zu rechnen. Die bereits zitierten Simulationsstudien haben zudem gezeigt, dass die Dimen-

sionalität der im Modell verwendeten Daten einen erheblichen Einfluss auf die Dichte am effizienten Rand hat. Mit steigender Zahl der Dimensionen nimmt also die Verzerrung und damit der ermittelte Effizienzwert zu.

Zum Vergleich der DEA-Ergebnisse von Stichproben unterschiedlicher Größen ist ein Verfahren zur Korrektur von Verzerrungen erforderlich, die durch unterschiedliche Stichprobengrößen in DEA-Modellen entstehen. Verschiedene Autoren haben hier ein Bootstrapping-Verfahren vorgeschlagen.<sup>45</sup>

Die Schätzung  $\hat{\theta}_k$  des wahren Effizienzwerts  $\theta_k$  von DMU  $k$  sowie die entsprechende Bootstrap-Schätzung  $\hat{\theta}_k^*$  stehen in folgender Beziehung zueinander:

$$(\hat{\theta}_k - \theta_k) | S \stackrel{approx.}{\sim} (\hat{\theta}_k^* - \hat{\theta}_k) | S^* \quad (9)$$

Wie im Rahmen von Bootstrap-Ansätzen üblich, wird die Beziehung des DEA-Schätzers zum wahren Effizienzwert (vgl. die linke Seite von (9)) durch die Beziehung des Bootstrap-Schätzers zum DEA-Schätzer approximiert (vgl. die rechte Seite des obigen Ausdrucks). Deshalb kann die Verzerrung der DEA-Effizienzwerte,  $bias_{S,k} = E_S(\hat{\theta}_k) - \theta_k$ , im allgemeinen Fall durch die mit \* gekennzeichnete Bootstrap-Entsprechung  $bias_{S^*,k} = E_{S^*}(\hat{\theta}_k^*) - \hat{\theta}_k$  approximiert werden. So kann ein verzerrungskorrigierter Schätzer  $\tilde{\theta}_k = \hat{\theta}_k - bias_{S^*,k} = 2\hat{\theta}_k - \hat{\theta}_k^*$  ermittelt werden. Hierbei bezeichnet  $S$  die jeweilige Stichprobe sowie  $\hat{\theta}_k^*$  das arithmetische Mittel der Effizienzwerte über alle DMUs:  $\hat{\theta}_k^* = B^{-1} \sum_B \hat{\theta}_k^*$ . Damit eine Verzerrungskorrektur in einer Verbesserung der Ergebnisse des gewöhnlichen DEA-Ansatzes resultiert, sollte die Korrektur allerdings nicht zu einem höheren mean square error (MSE) als bei der Ausgangsschätzung mit der Standardmethode führen. Hierzu muss die Bedingung  $bias / \hat{\sigma} > 3^{-1/2}$  erfüllt sein. Für den Fall, dass das MSE-Kriterium durch die Stichprobe nicht erfüllt wird, muss man schlussfolgern, dass die Daten stark von der Normalverteilung abweichen. In diesem Fall sind die verzerrungskorrigierten Effizienzwerte (Bootstrap-Effizienzwerte) nicht sinnvoll zu interpretieren, und es sollten die Ergebnisse des Standardmodells verwendet werden.<sup>46</sup>

Der Algorithmus zur Ableitung der um die Verzerrung durch unterschiedliche Stichprobengrößen korrigierten DEA-Schätzfunktion ist im Detail in Simar/Wilson (1998, Abschnitte 4.1 und 4.2) beschrieben. An dieser Stelle werden lediglich die

43 Vgl. Vanden Eeckaut/Tulkens/Jamar (1993), S. 308.

44 Simar/Wilson (2000), S. 59.

45 Vgl. Simar/Wilson (1999, 2000); Kneip/Simar/Wilson (2003).

46 Vgl. Gijbels et al. (1999).

Eckpunkte der Methode dargestellt. Zunächst wird, wie bereits in Abschnitt 3.2. beschrieben, angenommen, dass sich der Prozess der Erzeugung der Effizienzwerte  $\theta_k$  (der sog. data generating process) mit  $(\theta_1, \dots, \theta_n) \sim \text{i.i.d. } F(x, y)$  beschreiben lässt, wobei  $F(x, y)$  eine Dichtefunktion über  $(0, 1]$  darstellt. Der Prozess der Erzeugung von  $x_k$  (bei gegebenen, beobachteten Outputwerten  $y_k$  und Effizienzwert  $\theta_k$ , vgl. Formel (3)) wird somit vollständig durch die Dichtefunktion beschrieben. Von der Dichtefunktion  $\hat{F}(x, y)$  über  $\theta$  werden nun Bootstrap-Stichproben  $(x_k, y_k)_{k=1, \dots, n}$  gezogen, wobei  $\hat{F}(x, y)$  ein Schätzer der gemeinsamen Dichte von  $(x, y)$  über  $\theta$  bezeichnet. Da die inputorientierten Effizienzwerte von oben durch 1 beschränkt sind, weist  $\hat{F}(x, y)$  eine »Masse« von effizienten Beobachtungen mit  $\hat{\theta}_k = 1$  aus. Folglich wird die Schätzung der Dichte der Effizienzwerte in der Nähe der oberen Schranke (1) stark verzerrt sein. Eine Möglichkeit, diese Beschränkung nach oben zu berücksichtigen, besteht darin, die empirische Dichtefunktion  $\hat{F}(x, y)$  durch einen sog. »Kernel Smoother« zu glätten. Der optimale Wert für den Glättungsparameter (die sog. Bandbreite des Kernel-Schätzers) ist der Wert, der die mittlere quadratische Abweichung (den mean square error – MSE) minimiert. Die Kenntnis dieser Bandbreite genügt für die Simulation von Effizienzwerten aus der entsprechenden Verteilung. Auf Basis der optimalen Bandbreite und zufällig gezogener, entsprechend verteilter Störgrößen können durch Ziehen mit Zurücklegen aus Effizienzwerten der Ausgangsschätzung sog. Pseudo-Effizienzwerte erzeugt werden, die der Verteilung der Effizienzwerte aus der Ausgangsschätzung folgen. Hierbei werden die Störgrößen zu Effizienzwerten addiert, wobei sichergestellt wird, dass die Pseudo-Effizienzwerte innerhalb des Wertebereichs der Verteilung der Originalwerte liegen. Bei Kenntnis dieser Effizienzwerte können dann – im inputorientierten Fall – für bestimmte Outputwerte jeweils die zugehörigen Inputs berechnet werden. Auf Basis dieser neuen Input- und Outputdaten (sog. Pseudo-Daten) werden nun durch ein erneutes Durchführen einer DEA die Bootstrap-Effizienzwerte berechnet. Dieses Verfahren wird  $B$  mal wiederholt, wobei  $B$  »hinreichend groß« gewählt werden muss. Schließlich kann die Verzerrung des Schätzers auf Basis der o. a. Formeln korrigiert werden.

Diese Korrektur eröffnet zudem die Möglichkeit, bestimmte Hypothesentests in Bezug auf DEA-Modelle und ihre Spezifikationen durchzuführen. Ins-

besondere ist es möglich, Tests auf Effizienzunterschiede hinsichtlich bestimmter Teilstichproben durchzuführen.

#### *Empirische Illustration*

Es gibt eine lange Tradition von empirischen Studien, die sich mit der Messung von Markteffizienz bzw. Marktineffizienz befassen. Die empirische Literatur zur Effizienz von Konsumgütermärkten hat sich zur Bewertung der Markteffizienz anfangs auf das Vergleichen von Preis-Qualitäts-Korrelationen unterschiedlicher Produkte konzentriert. Diese Methodik erfordert es, die Qualitätsdimensionen eines Produktes in einem einzigen Index zu aggregieren. Daher wurde in den vergangenen zwei Jahrzehnten verstärkt die DEA als Methode zur Bestimmung der Markteffizienz herangezogen, da sie die Bestimmung der relativen Effizienz eines Produktes bei simultaner Berücksichtigung multidimensionaler Qualitätsaspekte ermöglicht.<sup>47</sup>

In einer richtungsweisenden Studie untersuchen Kamakura/Ratchford/Agrawal (1988) 20 Märkte mittels DEA. Auf jedem dieser Märkte werden zwischen 18 und 47 Produkte beobachtet; diese Produkte werden anhand von 2 bis 10 Eigenschaften charakterisiert. Die Autoren stellen fest, dass 52% aller Produkte ineffizient sind, und dass die durchschnittliche Ineffizienz bei 10% liegt. Ineffizienz eines Produktes von 10% bedeutet dabei, dass es ein Referenzprodukt auf dem Markt gibt, welches dem Konsumenten die gleichen Ausprägungen an Leistungseigenschaften (Output) für einen 10% geringeren Preis (Input) bietet. Eine spätere auf der DEA basierende Arbeit von Ratchford et al. (1996) beinhaltet 60 Märkte mit durchschnittlich 17 Produkten. Die Resultate ergeben eine durchschnittliche Ineffizienz von 18%. Daneben gibt es eine Reihe weiterer DEA-basierter Analysen, in denen jeweils ein einzelner Markt im Detail untersucht wird. Die Resultate dieser Arbeiten implizieren häufig nicht unbeträchtliche Marktineffizienzen.<sup>48</sup>

Es ist offensichtlich von Interesse, die Resultate verschiedener Studien zu vergleichen. Wie oben bereits dargestellt, sind aufgrund der Verzerrung des DEA-Schätzers die Effizienzen unterschiedlicher Produktmärkte nicht mehr vergleichbar, sobald diese sich aus einer unterschiedlichen Zahl

47 Vgl. Hammerschmidt (2006); Staat/Bauer/Hammerschmidt (2002).

48 Vgl. Chumpitaz/Kerstens/Paparoïdamis (2006).

Produkt	Beobachtungen (1)	Variablen (2)	BB (3)	Standard-DEA (4)	Bootstrap (bias-korrigiert) (5)	(4) – (5)	MSE verfehlt
Notebooks	18	6	0,05	96,84 %	93,01 %	3,96 %	18
Scanner	26	6	0,05	93,60 %	86,75 %	7,32 %	4
Drucker	33	6	0,07	91,21 %	81,42 %	10,73 %	20
Monitore 17"	17	6	0,11	88,83 %	77,35 %	12,91 %	0
PCs	13	6	0,12	89,42 %	76,04 %	14,97 %	13

Tab. 2: Verzerrungskorrigierte Effizienzwerte für 5 Produktmärkte

von Produkten (DMUs) zusammensetzen. Zudem könnte die in der Literatur oft geäußerte Feststellung, dass Märkte für Konsumgüter in westlichen Volkswirtschaften im Allgemeinen funktional effizient sind, auf die Eigenschaften des DEA-Schätzers und das unterschiedliche Ausmaß der Verzerrung aufgrund unterschiedlicher Stichprobengrößen zurückzuführen sein.

Um dies zu prüfen, wird im Folgenden die Markteffizienz für mehrere Märkte für Computer und Computerzubehör geschätzt, da die Originaldaten der oben zitierten Studien nicht verfügbar sind. Folgende Produktmärkte werden betrachtet (in Klammern die Anzahl der Beobachtungen):<sup>49</sup> Desktop PCs (13), 17"-Monitore (17), Notebooks (18), Scanner (26) und Tintenstrahldrucker (33). Als Input wird der Preis, als Outputs werden die Eigenschaften Qualität, Ausstattung mit Zusatzfunktionen, Serviceumfang, Verständlichkeit der Gebrauchsanweisungen sowie Ergonomie herangezogen. Die Daten wurden in der deutschen Computerzeitschrift »CHIP« in den Jahren 2001 und 2002 veröffentlicht. Alle Produkte wurden von CHIP mit demselben Versuchsaufbau getestet. Aus der obigen Diskussion über die Eigenschaften des DEA-Schätzers ging hervor, dass aufgrund der unterschiedlichen Stichprobengrößen unterschiedliche Verzerrungen der durchschnittlichen Effizienzwerte zu erwarten sind. Ohne eine Verzerrungskorrektur kann somit nicht gesagt werden, ob Unterschiede in den DEA-Effizienzwerten zwischen den Produktmärkten durch Verzerrungen bedingt sind oder tatsächliche Effizienzunterschiede repräsentieren. Wird keine Verzerrungskorrektur für diese Märkte

vorgenommen, würden sich die Effizienzwerte selbst dann unterscheiden, wenn die wahre durchschnittliche Effizienz für diese Märkte exakt dieselbe wäre.

Die Berechnung der Effizienz der Produkte basiert wie bei Kamakura/Ratchford/Agrawal (1988) auf einem inputorientierten Modell mit variablen Skalenerträgen (d. h.  $\zeta$  = variable returns to scale = VRS), wobei die Produkteigenschaften als Outputs fungieren, während der Input durch den Produktpreis gegeben ist. Um das Ausmaß der Verzerrung des Effizienzmaßes für die Produktmärkte abzuschätzen, wird das oben dargestellte Bootstrap-Verfahren angewendet. Tabelle 2 enthält Informationen über die Datensätze sowie die Ergebnisse. In den ersten beiden Datenspalten sind die Anzahl der Beobachtungen und die Anzahl der Input- und Outputvariablen aufgeführt. Die dritte Spalte zeigt die Bandbreite (BB), die für das Bootstrap-Verfahren verwendet wurde. Als nächstes werden die Standard-DEA- und die Bootstrap-Ergebnisse sowie die Differenz zwischen den beiden aufgeführt.

Die letzte Spalte von Tabelle 2 zeigt die Anzahl von Beobachtungen, für die die Bootstrap-Ergebnisse das MSE-Kriterium nicht erfüllen. Bei hoher Anzahl ist es angebracht, die Standard-DEA-Ergebnisse als Indikatoren der Effizienz zu verwenden und auf eine Verzerrungskorrektur zu verzichten. Für den Markt der 17"-Monitore ist das MSE-Kriterium für alle Beobachtungen, für Scanner für einen erheblichen Teil der Beobachtungen erfüllt.

Insbesondere für diese beiden Märkte ist es daher sinnvoll, die durchschnittliche, anhand des Standard-Modells ermittelte Marktineffizienz den bias-korrigierten Ergebnissen gegenüberzustellen. Wie ersichtlich ist, wird ohne Verzerrungskorrektur

<sup>49</sup> Vgl. im Folgenden Staat (2006).

die Effizienz der Märkte deutlich überschätzt. So wird für den Markt der 17-Monitore durch das Standardmodell nur etwa die Hälfte der tatsächlichen (d. h. korrigierten) Marktineffizienz ausgewiesen (11,17% im Standardmodell vs. 22,65% im Bootstrap-Modell). Es ist aufgrund des Problems der Verzerrung davon auszugehen, dass die Ergebnisse von auf Standard-DEA-Modellen basierenden Arbeiten die Markteffizienz in der Regel unrealistisch widerspiegeln.

#### 4. Zusammenfassung und Ausblick auf zukünftige Forschungsbereiche

Im Rahmen des vorliegenden Beitrages sind die grundsätzlichen Probleme der Datenqualität in Effizienzanalysen auf Basis der Data Envelopment Analysis (DEA) vorgestellt worden. Ausgangspunkt der Überlegungen war die Beobachtung, dass die DEA eine weit verbreitete Methode der relativen Effizienzmessung darstellt und bereits in vielen unterschiedlichen Bereichen ihre methodischen Stärken offenbaren konnte. Ein Blick in die mittlerweile umfangreiche Literatur zur DEA erbrachte den Schluss, dass Ex-post-Analysen zur (Güte-)Beurteilung von DEA-Ergebnissen vergleichsweise häufig diskutiert werden. Allerdings sind die grundsätzlichen Aspekte der Datenqualität – nämlich diejenigen, die sich auf die Festlegung der für eine DEA benötigten Datenbasis beziehen – bisher nur am Rande betrachtet wurden.<sup>50</sup> Hier bzw. in der Darstellung besonders geeigneter (Lösungs-)Ansätze lag der Fokus des Beitrages.

Das erste grundlegende Problem der Datenqualität – die Wahl der Vergleichseinheiten – ließ sich durch das Problem der Ausreißeridentifikation und -eliminierung beschreiben. Das Order-*m*-Verfahren konnte hier als »fortschrittlicher« Lösungsansatz identifiziert werden. Am Beispiel der Effizienzanalyse von Händlern eines deutschen Automobilunternehmens wurde die Anwendung dieses Verfahrens demonstriert. Im Rahmen des zweiten Problemkreises – der Festlegung der Effizienzkriterien – fanden zwei eng miteinander verknüpfte Aspekte Berücksichtigung: Zum einen muss der Anwender einer DEA entscheiden, welche für die Effizienz relevanten Merkmale grundsätzlich von den DMUs beeinflusst werden können. Dies führt zu einem DEA-Modell, das zwischen diskretionären und nicht diskretionären Variablen unterscheidet. Zum

anderen ist die Frage zu klären, ob die einzelnen Merkmale tatsächlich die Höhe der Effizienzwerte beeinflussen. Hierzu können Signifikanztests zum Einsatz gelangen. Als methodisch ausgereiftester Lösungsansatz im Rahmen dieser Probleme konnte der zweistufige Bootstrapping-Ansatz nach Simar/Wilson (2007) identifiziert werden, der anhand eines Beispiels der Effizienzanalyse von Bankfilialen illustriert wurde. Es zeigte sich, dass ohne diesen Ansatz ein fehlspezifiziertes Modell resultiert, dessen Anwendung möglicherweise zu irreleitenden Ergebnissen führt. Beim dritten, abschließend betrachteten Problem ging es um die Dimensionen der Datenmatrix einer DEA insgesamt. Hier ist ebenfalls ein Bootstrapping-basierter Ansatz sinnvoll, um stichprobenbedingte Verzerrungen zu korrigieren. Dies zeigte sich auch am empirischen Beispiel der Untersuchung zur Marktineffizienz.

Allerdings sei darauf hingewiesen, dass die hier dargestellten Verfahren zur Behandlung grundlegender Datenqualitätsaspekte in einer DEA erstens auf Basis von qualitativen Überlegungen ausgewählt worden sind und zweitens nicht frei von jeder Kritik sind. In diesem Sinne besteht jeweils weiterer Forschungsbedarf. Bezüglich der dargestellten Bootstrapping-Ansätze ist festzustellen, dass diese derzeit zwar die einzige Methode darstellen, um eine konsistente Schätzung für die Verteilung der Effizienzwerte zu erhalten, empirische Bewährungen bzw. Anwendungen bisher aber kaum existieren.<sup>51</sup> Zudem wären weitere Simulationsstudien erforderlich, die die generelle Vorteilhaftigkeit gegenüber herkömmlichen Ad-hoc-Ansätzen auch für weitere Datenkonstellationen belegen. Zudem ist zu betonen, dass bei sehr großen Stichproben ein Bootstrapping-Verfahren tendenziell nicht notwendig ist, da die Verzerrung der Effizienzwerte – bei gleichbleibender Zahl der Effizienzkriterien – mit zunehmender Stichprobengröße abnimmt.<sup>52</sup> Je kleiner die Zahl der Beobachtungen ist, desto höher ist der Nutzen eines Bootstrapping für praktische Anwendungen. Dies gilt gleichermaßen für den Order-*m*-Schätzer.<sup>53</sup>

50 Vgl. Pedraja-Chaporro/Salinas-Jimenez/Smith (1999) und Dyson et al. (2001) als nennenswerte Ausnahmen.

51 Zu den wenigen existierenden Anwendungen gehören aktuell Hammerschmidt/Falk/Bauer (2009) aus dem Marketing sowie Alexander/Haug/Jaforullah (2008), Barros/Dieke (2008) und de Witte/Marques (2008) außerhalb des Marketing-Bereiches.

52 Vgl. Kneip/Simar/Wilson (2003).

53 Vgl. Banker/Chang (2006).

Des Weiteren gibt es über die dargestellten Ansatzpunkte zur Sicherung der Datenqualität in einer DEA – nämlich eine angemessene Auswahl der Vergleichseinheiten (im Sinne der Ausreißeridentifikation) und der Effizienzkriterien (im Sinne der Trennung in diskretionäre und nicht diskretionäre Variablen sowie im Sinne einer Signifikanzprüfung derselben) – weitere Aspekte, die jedoch erst dann relevant werden, wenn die von uns genannten grundlegenden Probleme gelöst worden sind. Zudem existiert zu diesen nachgelagerten Problemen bereits eine Art »Leitfaden«.<sup>54</sup> Als Beispiele hierfür seien die Themenkreise »Operationalisierung und Messung der Effizienzkriterien« sowie »Fehlende Werte bei einzelnen DMUs« angeführt. Das erstgenannte Thema adressiert potenzielle Probleme, die durch die Verwendung unterschiedlicher Maßeinheiten für die Effizienzkriterien entstehen, insbesondere bei Vermischung von absoluten und relativen Größen wie z. B. Indizes. Im Rahmen der Vorgehensweise bei fehlenden Werten wird derzeit der Ansatz einer sog. *Fuzzy DEA* diskutiert.<sup>55</sup>

Abschließend sei angemerkt, dass positive Beurteilungen der DEA als »a powerful technique to assist managers«<sup>56</sup> oder »a widely used method to identify and adopt best practices as a means to improve performance and increase productivity«<sup>57</sup> zwar durchaus gerechtfertigt sind. Allerdings darf nicht vergessen werden, dass der Nutzen der DEA nur dann zur vollen Entfaltung kommen (d. h. sinnvolle und realistische Effizienzinformationen liefern) kann, wenn der Anwender die grundsätzlichen Schritte zur Sicherstellung von Datenqualität durchläuft, die wir hier präsentiert und illustriert haben. Diese werden allzu häufig gar nicht oder bestenfalls en passant in einem Nebensatz abgehandelt. Die Konzeption des Beitrages als kombinierte Darstellung geeigneter Lösungsansätze und deren Veranschaulichung an realen Datensätzen zielt letztlich darauf ab, die Anwendung dieser Ansätze nachvollziehbar zu machen und damit auch ihre Verbreitung zu fördern.

54 Vgl. Dyson et al. (2001).

55 Vgl. Smirlis/Maragos/Despotis (2006) und Maragos/Despotis (2004).

56 Donthu/Hershberger/Osmonbekov (2005), S. 1481.

57 Zhu (2003), S. xxi.

## Verzeichnis der zitierten Literatur

- Alexander, W. R. J./Haug, A. A./Jaforullah, M. (2008): A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools. *Economics Discussion Papers No. 0714*, University of Otago 2008.
- Allen, K. (2002): Messung ökologischer Effizienz mittels Data Envelopment Analysis. Wiesbaden 2002.
- Banker, R. D./Chang, H. (2006): The Super-Efficiency Procedure is for Outlier Identification, Not for Ranking Efficient Units. In: *European Journal of Operational Research*, Vol. 175 (2006), Nr. 2, S. 1311–1320.
- Banker, R. D./Das, S./Datar, S. M. (1989): Analysis of cost variances for management control in hospitals. In: *Research in Governmental and Nonprofit Accounting*, Vol. 5 (1989), S. 269–291.
- Banker, R. D./Gifford, J. L. (1988): A relative efficiency model for the evaluation of public health nurse productivity. Mellon University Mimeo, Carnegie 1988.
- Barros, C. P./Dieke, P. U. C. (2008): Measuring the economic efficiency of airports: A Simar–Wilson methodology analysis. In: *Transportation Research: Part E*, Vol. 44 (2008), Nr. 6, S. 1039–1051.
- Bauer, H. H./Staat, M./Hammerschmidt, M. (Hrsg.): *Marketingeffizienz: Messung und Steuerung mit der DEA – Konzept und Einsatz in der Praxis*, München 2006.
- Cazals, C./Florens, J.-P./Simar, L. (2002): Nonparametric frontier estimation: A robust approach. In: *Journal of Econometrics*, Vol. 106 (2002), Nr. 1, S. 1–25.
- Charnes, A./Cooper, W. W./Rhodes, E. (1978): Measuring the efficiency of decision making units. In: *European Journal of Operational Research*, Vol. 2 (1978), Nr. 6, S. 429–444.
- Cherchye, L./Post, T. (2003): Methodological Advances in DEA: A survey and an application for the Dutch electricity sector. In: *Statistica Neerlandica*, Vol. 57 (2003), Nr. 4, S. 410–438.
- Chumpitaz, R./Kerstens, K./Paparoidamis, N. (2006): Marketing-performance: Von ein- zu multidimensionalen und von Durchschnitts- zu Frontier-Konzepten. In: Bauer, H. H./Staat, M./Hammerschmidt, M. (Hrsg.): *Marketingeffizienz: Messung und Steuerung mit der DEA – Konzept und Einsatz in der Praxis*, München 2006, S. 3–30.
- Cooper, W. W./Huang, Z. M./Li, S. X. (1996): Satisficing DEA models under chance constraints. In: *Annals of Operations Research*, Vol. 66 (1996), Nr. 4, S. 279–295.
- Cooper, W. W./Seiford, L. M./Tone, K. (2006): *Introduction to Data Envelopment Analysis and Its Uses*. New York 2006.
- Donthu, N./Hershberger, E. K./Osmonbekov, T. (2005): Benchmarking marketing productivity using data envelopment analysis. In: *Journal of Business Research*, Vol. 58 (2005), S. 1474–1482.
- Doyle, J. R./Green, R. H. (1994): Efficiency and Cross-Efficiency in DEA: Derivations, Meanings and Uses. In: *The Journal of the Operational Research Society*, Vol. 45 (1994), Nr. 5, S. 567–578.
- Dyckhoff, H. (1993): *Aktivitätsanalyse*. In: Wittmann, W. et al. (Hrsg.): *Handwörterbuch der Betriebswirtschaft*. 5. Aufl. Stuttgart 1993, Sp. 57–68.
- Dyckhoff, H./Allen, K. (1999): Theoretische Begründung einer Effizienzanalyse mittels Data Envelopment Analysis. In: *Zeitschrift für Betriebswirtschaftliche Forschung*, 51. Jg (1999), Nr. 5, S. 411–436.
- Dyckhoff, H./Allen, K. (2001): Measuring ecological efficiency with data envelopment analysis. In: *European Journal of Operational Research*, Vol. 132 (2001), Nr. 2, S. 312–325.
- Dyckhoff, H./Gilles, R. (2004): Messung der Effektivität und Effizienz produktiver Einheiten. In: *Zeitschrift für Betriebswirtschaft*, 74. Jg (2004), S. 765–783.

- Dyson, R. G./Allen, R./Camanho, A. S./Podinovski, V. V./Sarrico, C. S./Shale, E. A. (2001): Pitfalls and protocols in DEA. In: *European Journal of Operational Research*, Vol. 132 (2001), Nr. 2, S. 245–259.
- Gijbels, I. E./Mammen, E./Park, U./Simar, L. (1999): On Estimation of Monotone and Concave Frontier Functions. In: *Journal of the American Statistical Association*, Vol. 94 (1999), S. 220–228.
- Glaser, B./Kleine, A. (2004): Distance Maximization in DEA. Discussion Paper Operations Research, Universität Hohenheim 2004.
- Hammerschmidt, M. (2006): Effizienzanalyse im Marketing – Ein produktionstheoretisch fundierter Ansatz auf Basis von Frontier Functions. Wiesbaden 2006.
- Hammerschmidt, M./Falk, T./Bauer, H. H. (2009): Investigating the Brand Productivity Gap. Working paper, Universität Mannheim 2009.
- Hendricks, W./Koenker, R. (1992): Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. In: *Journal of the American Statistical Association*, Vol. 87 (1992), S. 58–68.
- Isik, I./Hassan, M. K. (2002): Technical, scale, and allocative efficiencies of Turkish banking industry. In: *Journal of Banking and Finance*, Vol. 26 (2002), S. 719–766.
- Kamakura, W. A./Ratchford, B. T./Agrawal, J. (1988): Measuring Market Efficiency and Welfare Loss. In: *The Journal of Consumer Research*, Vol. 15 (1988), Nr. 3, S. 289–302.
- Kleine, A. (2002): DEA-Effizienz – Entscheidungs- und produktions-theoretische Grundlagen der Data Envelopment Analysis. Wiesbaden 2002.
- Kneip, A./Simar, L./Wilson, P. W. (2003): Asymptotics for DEA Estimators in Nonparametric Frontier Models, Discussion paper #317, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve 2003.
- Koopmans, T. C. (1951): Analysis of Production as an Efficient Combination of Activities. In: Koopmans, T. C. (Hrsg.): *Activity Analysis of Production and Allocation*. New York 1951, S. 33–97.
- Löber, G.-A./Staat, M. (2006): Marketingrelevante Weiterentwicklungen der DEA. In: Bauer, H. H./Staat, M./Hammerschmidt, M. (Hrsg.): *Marketingeffizienz: Messung und Steuerung mit der DEA – Konzept und Einsatz in der Praxis*. München 2006, S. 61–105.
- Maragos, E. K./Despotis, D. K. (2004): The evaluation of the efficiency with data envelopment analysis in case of missing values: A fuzzy approach. In: *WSEAS Transactions on Mathematics*, Vol. 3 (2004), Nr. 3, S. 656–663.
- Mukherjee, K./Ray, S./Miller, S. M. (2001): Productivity growth in large US commercial banks: the initial post-deregulation experience. In: *Journal of Banking and Finance*, Vol. 25 (2001), S. 913–939.
- Paradi, J. C./Vela, S./Yang, Z. (2004): Assessing bank and bank branch performance: modelling considerations and approaches. In: Cooper, W. W./Seiford, L. M./Zhu, J. (Hrsg.): *Handbook on Data Envelopment Analysis*. Boston 2004, S. 349–400.
- Pareto, V. (1897): *Cours d'Economie Politique*. Lausanne/Paris 1897.
- Pedraja-Chaparro, F./Salinas-Jimenez, J./Smith, P. (1999): On the Quality of the Data Envelopment Analysis Model. In: *The Journal of the Operational Research Society*, Vol. 50 (1999), Nr. 6, S. 636–644.
- Ratchford, B. T./Agrawal, J./Grimm, P. E./Srinivasan, N. (1996): Toward Understanding the Measurement of Market Efficiency. In: *Journal of Public Policy & Marketing*, Vol. 15 (1996), Nr. 2, S. 167–184.
- Seiford, L. M. (2006): A Cyber-Bibliography for Data Envelopment Analysis (1978–2005). In: Cooper, W. W./Seiford, L. M./Tone, K. (Hrsg.): *Introduction to Data Envelopment Analysis and Its Uses*. New York 2006.
- Simar, L. (2003): Detecting Outliers in Frontier Models: A Simple Approach. In: *Journal of Productivity Analysis*, Vol. 20 (2003), Nr. 3, S. 391–424.
- Simar, L./Wilson, P. W. (1998): Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. In: *Management Science*, Vol. 44 (1998), S. 49–61.
- Simar, L./Wilson, P. W. (1999): Of Course We Can Bootstrap DEA Scores! But Does It Mean Anything? Logic Trumps Wishful Thinking. In: *Journal of Productivity Analysis*, Vol. 11 (1999), Nr. 1, S. 93–97.
- Simar, L./Wilson, P. W. (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art. In: *Journal of Productivity Analysis*, Vol. 13 (2000), Nr. 1, S. 49–78.
- Simar, L./Wilson, P. W. (2007): Estimation and inference in two-stage, semi-parametric models of production processes. In: *Journal of Econometrics*, Vol. 136 (2007), S. 31–64.
- Smirlis, Y. G./Maragos, E. K./Despotis, D. K. (2006): Data envelopment analysis with missing values: An interval DEA approach. In: *Applied Mathematics and Computation*, Vol. 177 (2006), S. 1–10.
- Staat, M. (1999): Treating non-discretionary variables one way or the other: implications for efficiency scores and their interpretation. In: Westermann, G. (Hrsg.): *Data Envelopment Analysis in the Service Sector*. Wiesbaden 1999, S. 23–49.
- Staat, M. (2006): Messung der Effizienz von Produktmärkten. In: Bauer, H. H./Staat, M./Hammerschmidt, M. (Hrsg.): *Marketingeffizienz: Messung und Steuerung mit der DEA – Konzept und Einsatz in der Praxis*. München 2006, S. 261–272.
- Staat, M./Bauer, H. H./Hammerschmidt, M. (2002): Structuring Product-Markets: An Approach Based on Customer Value. In: *AMA Winter Marketing Educators' Conference Proceedings*, Vol. 13 (2002), S. 205–212.
- Thanassoulis, E. (1999): Setting Achievement Targets for School Children. In: *Education Economics*, Vol. 7, Nr. 2, S. 101–119.
- Tavares, G. (2002): A bibliography of data envelopment analysis (1978–2001). *Rutcor Research Report* 2002.
- Vanden Eeckaut, P./Tulkens, P. H./Jamar, M. A. (1993): Cost efficiency in Belgian municipalities. In: Fried, H. O./Lovell, C. A. K./Schmidt, S. S. (eds.): *The measurement of productive efficiency*. Oxford/New York 1993, S. 300–334.
- Wilken, R. (2007): *Dynamisches Benchmarking. Ein Verfahren auf Basis der Data Envelopment Analysis*. Wiesbaden 2007.
- Wilson, P. W. (1995): Detecting influential observations in data envelopment analysis. In: *The Journal of Productivity Analysis*, Vol. 6 (1995), S. 27–45.
- de Witte, K./Marques, R. C. (2008): Designing incentives in local public utilities. In: Cabrera, E./Pardo, M. (Hrsg.): *Performance Assessment of Urban Infrastructure Services*. London 2008, S. 297–307.
- Zhang, Y./Bartels, R. (1998): The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand. In: *Journal of Productivity Analysis*, Vol. 9 (1998), Nr. 3, S. 187–204.
- Zhu, J. (2003): *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets and DEA Excel Solver*. Boston 2003.
- Zhu, J./Cook, W. D. (2007): *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*. New York 2007.