

## 2 Historie und aktueller Stand der Forschung in der DEA

Dieses Kapitel soll dem Leser als Einführung in die dieser Arbeit zu Grunde liegenden Bereiche der Wissenschaft dienen und die Grundlage für die weiteren theoretischen und empirischen Überlegungen bilden. Der Autor arbeitet dabei den aktuellen Stand der wissenschaftlichen Forschung auf und gibt ihn mit den für diese Arbeit relevanten Schwerpunkten wieder. Dabei wird auch eine mathematische Notation eingeführt, die im Sinne der Durchgängigkeit und Transparenz über die gesamte Arbeit Geltung haben wird.

### 2.1 Grundlagen der Effizienzmessung

#### 2.1.1 Allgemeines zum Begriff der Effizienz

Der Definition und anschaulichen Konkretisierung des Effizienzbegriffes kommt an dieser Stelle eine wesentliche Bedeutung zu – zum einen weil die Frage nach der sinnvollen Messung von Effizienz sowie deren angemessener Interpretation in dieser Arbeit das zentrale Leitmotiv darstellt. Zum anderen weil festgestellt werden muss, dass der Begriff der Effizienz sowohl im allgemeinen Sprachgebrauch als auch in weiten Teilen der einschlägigen Literatur nicht immer scharf umrissen ist und implizit gleichgesetzt wird mit Begriffen wie „Produktivität“, „Effektivität“, „Wirtschaftlichkeit“, „Leistungsfähigkeit“ und vielen mehr [BOH1993, S. 855]. Zunächst gilt es also, bezüglich dieser verschiedenen Begriffe Klarheit zu schaffen und eine definitorische Basis für die weiteren Ausführungen einzuführen. Um nicht lediglich die etablierten Begriffsdefinitionen nebeneinander zu stellen, soll hier von der Frage ausgegangen werden, zu welchem Zweck der Effizienzbegriff geschaffen wurde und über welche Fragestellungen er Aufschluss geben soll:

Die **Wohlfahrtstheorie** befasst sich mit der Beurteilung und Beeinflussung der gesellschaftlichen Wohlfahrt, also auf makroökonomischer Ebene. Dabei stellt sie die Frage nach einer optimalen Allokation von Ressourcen zum Wohle der

Gesellschaft, dem sogenannten Wohlfahrtsoptimum. Die Knappheit von (materiellen und immateriellen) Ressourcen wird dabei unterstellt [GAB2112]. Die wichtigste immaterielle Ressource ist die Zeit, die durch die Endlichkeit des Lebens naturgemäß für jeden Menschen ein nur begrenzt zur Verfügung stehendes Gut darstellt. Die materiellen Güter müssen hergestellt, veredelt oder wenigstens transportiert werden, bevor sie ihrem Adressaten den gewünschten Nutzen stiften – auch hier ist von einer Knappheit auszugehen. Die Beurteilung der gesellschaftlichen Wohlfahrt beruht auf rein subjektiven Nutzeneinschätzungen der Individuen [KUE1975, S. 1] [SOH1992, S. 1] [BRU2007, S. 42], denn es obliegt allein den Einzelpersonen, die Maßstäbe für die eigene Zufriedenheit bzw. Wohlfahrt festzulegen. Der Wunsch des Menschen nach immer besseren Lebensbedingungen und immer mehr Konsum ist jedoch unbestritten und auch empirisch zu beobachten. Insgesamt darf also davon ausgegangen werden, dass die Gesellschaft und ihre Individuen danach streben, mit möglichst geringem Ressourceneinsatz einen möglichst großen Nutzen zu erzielen.

Aus dieser Ausgangssituation leitet sich zwangsläufig die Frage ab, wie man ein angestrebtes Resultat mit möglichst geringem Gütereinsatz erreichen und/oder mit gegebenen Mitteln ein maximales Ergebnis erwirtschaften kann [MEW1985, S. 262]. Es geht also um die Relation zwischen eingebrachtem Einsatz und erzieltm Ergebnis. Wenn man nun den Einsatz als **Input** und das Ergebnis als **Output** bezeichnet, lässt sich diese Relation, die als **Produktivität** bezeichnet wird, so darstellen:

$$\text{Produktivität} = \frac{\text{Output}}{\text{Input}}$$

Wie sich leicht erkennen lässt, handelt es sich bei der so definierten Produktivität um ein absolutes Maß, das ohne die Einbeziehung eines Bewertungsmaßstabs, z. B. einer Vergleichsproduktivität, kaum Aussagekraft entfalten kann. Hier kommt nun der Begriff der Effizienz ins Spiel: Die Produktivität ist zwar der wesentliche Kern einer jeden Effizienzbetrachtung, doch liegt der entscheidende Unterschied

darin, dass ein Effizienzmaß seine Aussagekraft dadurch gewinnt, eine Produktivität ins Verhältnis zu einer Vergleichsgröße zu setzen [PET2008, S. 707]. Bei der Effizienz handelt es sich also um ein relatives Maß [SCH2000, S. 3] [DYS2010, S. 116]. Allgemein ist die (Produktions-) **Effizienz** definiert als ein „Zustand, in dem es bei gegebener Ressourcenausstattung und Technologie nicht möglich ist, von mind. einem Gut mehr und von allen anderen Gütern mind. genauso viel herzustellen“ [GAB2012], auch als PARETO-Optimum bezeichnet. Die hier angeführte Begrenztheit der Ressourcen haben wir bereits begründet; der Begriff der Technologie und wie Vergleichsgrößen durch diese Technologie in das Effizienzmaß eingehen, wird im folgenden Abschnitt näher erläutert.

Häufig werden die Begriffe Effizienz und **Effektivität** fälschlicherweise als Äquivalente verwendet. Der Unterschied in der Bedeutung könnte jedoch größer nicht sein: Während die Effizienz das Verhältnis von In- und Output zueinander ins Zentrum setzt, handelt es sich bei der Effektivität um eine reine Outputbetrachtung. Hier wird nicht nach dem Aufwand gefragt, sondern ausschließlich nach dem erzielten Ergebnis, also dem Zielerreichungsgrad – der Ressourceneinsatz wird dabei völlig außen vor gelassen. So wird dem US-amerikanischen Ökonomen PETER F. DRUCKER folgender Ausspruch zugeschrieben: „... efficiency is about doing the things right and effectiveness is about doing the right things.“ [AJG1995, S. 333]. Pointiert fügt DRUCKER hinzu: “There is surely nothing quite so useless as doing with great efficiency what should not be done at all.” [DRU2006, S. 147]. Treffender kann man die beiden Begriffe kaum voneinander abgrenzen. Grundsätzlich ist auf Effizienz und Effektivität gleichermaßen Wert zu legen, weil das Eine nur in Kombination mit dem Anderen eine sinnvolle Wirkung entfalten kann: Effiziente Maßnahmen verlieren an Wirkung, wenn sie nicht exakt auf die angestrebten Ziele ausgerichtet sind. Umgekehrt kommen effektive Maßnahmen nur zu einer schnellen und ressourcenschonenden Zielerreichung, wenn sie gleichzeitig auch effizient sind.

## 2.1.2 Überblick über Entwicklung und Methoden der Effizienzmessung

### 2.1.2.1 Produktion und Technologiemenge

**Produktion** ist definiert als eine durch Menschen veranlasste Transformation von Inputs zu Outputs mit dem Ziel einer Nutzenerhöhung (Wertschöpfung). Diese Transformation, schematisch dargestellt in Abbildung 2.1, findet in der Regel in einer Wirtschafts- oder Produktionseinheit (DMU = Decision Making Unit) statt und muss nicht zwangsläufig eine qualitative Veränderung der Inputobjekte mit sich bringen [DYS2010, S. 3]. Im Jahr 2010 begannen zwei Geschäftsleute den Vertrieb von Glühlampen unter der Bezeichnung „heatball“, also nicht als Licht-, sondern als Wärmequelle, insbesondere zur Umgehung der EG-Verordnung Nr. 244/2009 zum Verbot von energieineffizienten Punktlichtquellen. Hier liegt die Wertschöpfung in der reinen Deklaration des Produktes, auch wenn der „heatball“ letztlich aufgrund verschiedener Gerichtsentscheidungen wieder aus dem Handel genommen werden musste [ZEO2012].



Abbildung 2.1: Transformation der Input-Faktoren zu Output-Faktoren  
(Quelle: eigene Darstellung)

In- und Output können sowohl materieller als auch immaterieller Art sein [COG2007, S. 110 ff.]. Als wichtigste immaterielle Inputs sind die in den Transformationsprozess eingebrachten Patente, Lizenzen und Rechte zu nennen, aber auch immaterielle Outputs sind keine Seltenheit. So leben beispielsweise Werbeagenturen in erster Linie von der Erschaffung kreativer Ideen für den Verkauf der Kundenprodukte.

Bevor nun der Begriff der Technologiemenge eingeführt werden kann, bedarf es der Festlegung einer mathematischen Notation:

Eine Gesamtmenge von  $J$  Wirtschaftseinheiten (**DMUs**)

$$DMU_j = \text{Decision Making Unit } j \text{ mit } j = 1, \dots, J$$

verfüge über  $M$  Produktionsfaktoren (**Inputs**)

$$x_{jm} = \text{Input } m \text{ der } DMU_j \text{ mit } \mathbf{x}_j = (x_{j1}, \dots, x_{jm}, \dots, x_{jM})^T \in \mathbb{R}_+^M$$

zur Erstellung von  $S$  Produkten (**Outputs**), bei denen es sich um End- oder Zwischenprodukte handeln kann:

$$y_{js} = \text{Output } s \text{ der } DMU_j \text{ mit } \mathbf{y}_j = (y_{j1}, \dots, y_{js}, \dots, y_{jS})^T \in \mathbb{R}_+^S$$

Bei den in der DMU stattfindenden Transformationsprozessen, also der Erstellung der gewünschten Produkte, können auch sogenannte Nebenprodukte entstehen, die dann entweder weiterverarbeitet, verkauft oder entsorgt werden.

Die jeweiligen (nichtnegativen) Quantitäten der  $M$  Input- und  $S$  Outputfaktoren werden für jede DMU mit einem **Inputvektor**  $\mathbf{x}$  und einem **Outputvektor**  $\mathbf{y}$  erfasst. Die Quantität wird dabei in den (unterschiedlichen) Maßeinheiten der einzelnen Faktoren ausgedrückt. Die Inputvektoren  $\mathbf{x}_j$  und  $\mathbf{y}_j$  der Wirtschaftseinheiten (DMUs)  $j$  sollen als zeitraumbezogene empirische Ausprägungen verstanden werden:

$\mathbf{x}_j$  = Inputvektor  $\mathbf{x}$  der  $DMU_j$  mit  $\mathbf{x}_j \in \mathbb{R}_+^M$

$\mathbf{y}_j$  = Outputvektor  $\mathbf{y}$  der  $DMU_j$  mit  $\mathbf{y}_j \in \mathbb{R}_+^S$

Durch die Kombination von Inputvektor  $\mathbf{x}_j$  und Outputvektor  $\mathbf{y}_j$  entsteht der Produktionsvektor bzw. die **Aktivität**

$(\mathbf{x}_j, \mathbf{y}_j)$  = Aktivität  $(\mathbf{x}, \mathbf{y})$  der  $DMU_j$  mit  $(\mathbf{x}_j, \mathbf{y}_j) \in \mathbb{R}_+^{M+S}$ .

Für  $J$  Produktionen existieren somit  $J$  Aktivitäten  $(\mathbf{x}_j, \mathbf{y}_j)$  in einem  $(M + S)$ -dimensionalen Vektorraum. Zur besseren Veranschaulichung sind in Abbildung 2.2 (siehe folgende Seite) die Aktivitäten  $(\mathbf{x}_j, \mathbf{y}_j)$  der Beispiel-DMUs 1, 2, 3 und 4 in einem zweidimensionalen Raum dargestellt. Die konkreten Aktivitäten dieser Beispiel-DMUs sind einem Diskussionsbeitrag von RÖDDER UND DELLNITZ [ROD2011] entnommen und werden zum Zweck einer höchstmöglichen Anschaulichkeit im weiteren Verlauf immer wieder für graphische Darstellungen herangezogen.

Allgemein ist die **Technologiemenge**  $\mathbf{T}$ , im Folgenden auch als Technologieraum bezeichnet, definiert als die Menge aller in einer Wirtschaftseinheit technisch möglichen In- und Outputkombinationen (Aktivitäten) [STE2012] und somit als der Raum der möglichen Produktionen. Dieser Technologieraum spielte bereits im Jahr 1951 in den Schriften von DEBREU [DEB1951] und KOOPMANS [KOO1951] eine Rolle – ihm liegt die Annahme zugrunde, dass allen betrachteten DMUs dieselbe Technologie zur Verfügung steht und somit jede beobachtete Aktivität auch für jede DMU theoretisch erreichbar ist. Der Technologieraum beschreibt die technischen und organisatorischen Fähigkeiten aller DMUs in ihrer Gesamtheit, während die darin liegenden Aktivitäten verdeutlichen, inwieweit jede einzelne DMU diese Möglichkeiten ausgeschöpft hat. Um dieser Aussage gerecht werden zu können, muss der Technologieraum ausschließlich über *vergleichbare* DMUs aufgespannt werden [SCH2000, S. 38]. Im folgenden Kapitel

wird diese Anforderung der Vergleichbarkeit der zu untersuchenden DMUs noch einmal aufgegriffen und präzisiert.

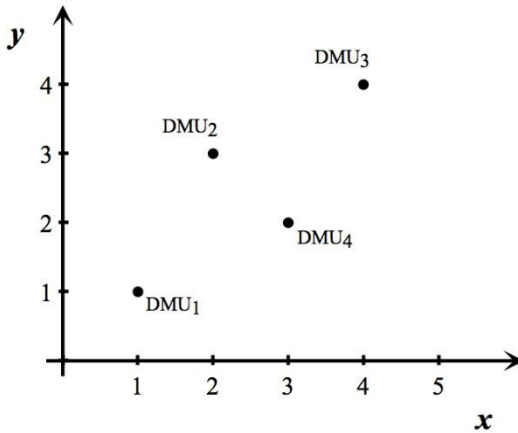


Abbildung 2.2: Aktivitäten der DMUs  $j = 1, \dots, 4$   
(Quelle: in Anlehnung an [ROD2011, S. 6])

Die **Produktionsfunktion** repräsentiert alle bestmöglichen und somit effizienten Produktionspunkte (Aktivitäten) auf Basis einer gegebenen Technologie und kann auch als funktionale Beschreibung des effizienten Randes der Technologiemenge betrachtet werden [CKH2007, S. 3 f.] [CRO2005, S. 3]. Abbildung 2.3 zeigt eine fiktive ertragsgesetzliche Produktionsfunktion anhand der bekannten Beispiel-DMUs.

Wäre diese Produktionsfunktion einer Technologie bekannt, so ließe sich für jede unter dieser Technologie arbeitende DMU die bestmögliche bzw. möglichst effiziente Input-/Output-Kombination problemlos rechnerisch bestimmen [RRL2008, S. 290], also der maximal mögliche Output bei gegebenem Input oder der minimal mögliche Input bei gegebenem Output.

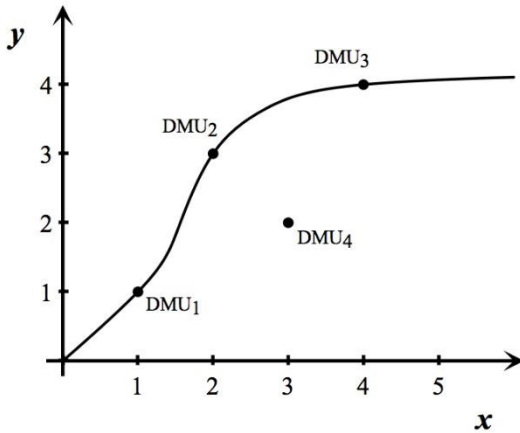


Abbildung 2.3: Beispielhafte ertragsgesetzliche Produktionsfunktion  
(Quelle: eigene Darstellung)

Im Fall einer nicht bekannten Produktionsfunktion, wie es im weiteren Verlauf dieser Arbeit unterstellt wird, kommt der Beobachtung der von den DMUs realisierten Aktivitäten besondere Bedeutung bei der Abschätzung der technisch möglichen In- und Outputkombinationen, also der Technologiemenge, zu. Eine erste intuitive Definition der Technologiemenge lautet

$$T = \{(x, y) \in \mathbb{R}_+^{M+S} \mid x \text{ kann } y \text{ produzieren}\}.$$

Hier wird bewusst auf den Index  $j$  verzichtet, da es sich bei „ $x$  kann  $y$  produzieren“ um eine allgemeine Aussage handelt, welche natürlich auch in Bezug auf jede DMU  $j$  Geltung hat. Außerdem sei darauf verwiesen, dass der Begriff „produzieren“ in diesem Zusammenhang nicht zu wörtlich verstanden werden sollte, da hier jede Art von Transformation angesprochen ist, also auch Dienstleistungen, Handel und vieles mehr.



Die Herleitung des Technologieraums aus den empirisch ermittelten Aktivitäten der DMUs wird als „nicht-parametrischer Ansatz“ bezeichnet und muss letztlich immer als eine Form der Schätzung betrachtet werden, da der tatsächliche, reale Technologieraum hier unbekannt ist. Über die beobachteten Aktivitäten der DMUs hinaus bestimmt sich die Definition der Technologiemenge  $T$  über Forderungen an ihre Struktur; dies geschieht in Form von Axiomen. Obwohl diese Axiome den meisten Lesern bekannt sind, bedarf es ihrer Einführung und Erläuterung, da sie vor dem Hintergrund verschiedener Anwendungen wieder aufgegriffen und diskutiert werden sollen:

### **(1) Vollständigkeit**

Wie vorab erwähnt, wird der Technologieraum  $T$  über die beobachteten Aktivitäten der DMUs  $J$  aufgespannt. Zwar wird die Technologiemenge dadurch nicht hinreichend beschrieben, da auch die hier angeführten Axiome zu berücksichtigen sind, doch müssen die Aktivitäten der betrachteten DMUs selbstverständlich Bestandteil der Technologiemenge sein [SEI2006, S. 5] [KLE2002, S. 130]. Es gilt

$$(x_j, y_j) \in T \text{ für alle } j = 1, \dots, J.$$

Diese Anforderung wird in der Literatur auch als „empirische Vollständigkeit“ bezeichnet.

### **(2) Möglichkeit von Ineffizienzen (bzw. Monotonie oder Verschwendbarkeit)**

Bei Unterstellung der Möglichkeit von Ineffizienzen kann der Input einer gegebenen Aktivität frei erhöht und/oder der Output gesenkt werden, ohne dass dadurch der Technologieraum verlassen würde:

Wenn  $(x, y) \in T$ ,  
dann ist auch  $(x', y) \in T$  bei  $x' \geq x$  und  $(x, y') \in T$  bei  $y' \leq y$

Praktisch ausgedrückt: Wenn ein Schuster ein handgefertigtes Paar Schuhe in 35 Arbeitsstunden herstellen kann, dann kann er das natürlich auch in 40 oder mehr Stunden tun, indem er langsamer arbeitet. Betrachtet man denselben Fall aus Output-Sicht, dann wird unter normalen Bedingungen bei einer 35-Stunden-Arbeitswoche genau ein Paar Schuhe pro Woche hergestellt. Natürlich kann der Output bei gegebener Arbeitsleistung (Input) sinken, zum Beispiel wenn der Schuster Fehler macht und dadurch Ausschuss produziert. Im angeführten Beispiel liegt eine sogenannte „freie Verschwendbarkeit“ vor, wie sie auch in diesem Axiom unterstellt wird. Das zeigt sich konkret darin, dass alle Inputs frei erhöht und alle Outputs frei, also unabhängig voneinander, reduziert werden können – bei nur einem In- und einem Output liegt das in der Natur der Sache. In Fällen mit mehreren In- und Outputs kann es jedoch vorkommen, dass sich die Inputs nur gemeinsam, also radial, erhöhen bzw. die Outputs nur radial reduzieren lassen, man spricht dabei von „schwacher Verschwendbarkeit“ [SCH2000, S. 45]. Grundsätzlich wird in dieser Arbeit freie Verschwendbarkeit unterstellt, in Unterkapitel 4.4 wird diese Thematik vor dem Hintergrund der DEA-Modellgestaltung intensiver beleuchtet.

Abbildung 2.4 zeigt einen Technologieraum auf Basis freier Verschwendbarkeit für die bekannten vier DMUs. Es lässt sich erkennen, dass, ausgehend von den Aktivitäten der DMUs, alle Steigerungen der jeweiligen Inputs (eng gestrichelte Pfeile) bis ins Unendliche sowie alle Senkungen der jeweiligen Outputs bis zur Menge von 0 (weit gestrichelte Pfeile) möglich sind und somit zum Bestandteil der Technologiemenge werden. Daraus und aus den daraus resultierenden Kombinationsmöglichkeiten ergibt sich der in Abbildung 2.4 dargestellte Technologieraum (grau hinterlegt), dessen „oberer“ Rand den Verlauf einer Treppenfunktion aufweist. Ein so definierter Technologieraum wird auch als „Free Disposal Hull“ bezeichnet und geht auf Arbeiten von DEPRINS ET AL. [DST1984] und TULKENS [TUL1993] zurück.

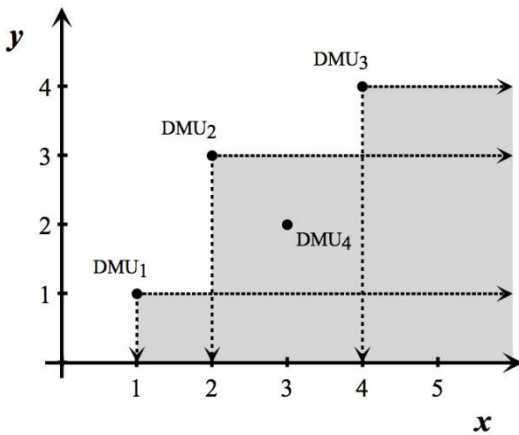


Abbildung 2.4: Technologieraum unter Annahme von Verschwendbarkeit  
(Quelle: eigene Darstellung)

### (3) *Konvexität*

Die Annahme der Konvexität geht davon aus, dass beliebige Aktivitäten derselben Technologie konvex kombiniert werden können:

$$(\mathbf{x}', \mathbf{y}') \text{ und } (\mathbf{x}'', \mathbf{y}'') \in T \text{ und } \lambda \in [0,1] \Rightarrow \lambda(\mathbf{x}', \mathbf{y}') + (1 - \lambda)(\mathbf{x}'', \mathbf{y}'') \in T$$

Graphisch gesehen werden durch das Axiom der Konvexität alle Verbindungslinien zwischen allen empirisch beobachteten und allen sonstigen (!) Aktivitäten zum Bestandteil der Technologiemenge. Abbildung 2.5 zeigt beispielhaft die beiden fiktiven Aktivitäten  $(\mathbf{x}', \mathbf{y}')$  und  $(\mathbf{x}'', \mathbf{y}'')$  sowie den Technologieraum, der durch die Summe der realen und der möglichen Aktivitäten aufgespannt wird:

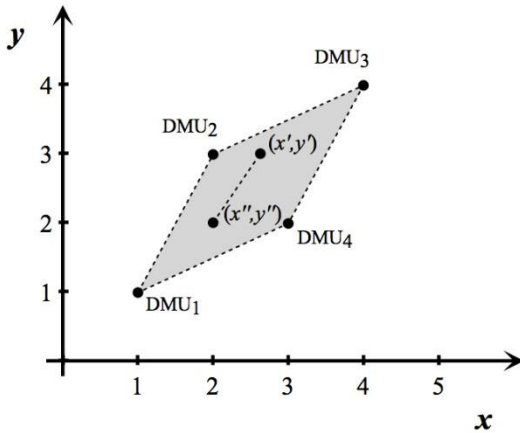


Abbildung 2.5: Graphische Auswirkungen der Konvexitätsannahme auf die Technologiemenge  
(Quelle: eigene Darstellung)

In Abbildung 2.6 findet sich die Darstellung eines Technologieraums bei Konvexitätsannahme und zusätzlich unterstellter freier Verschwendbarkeit. Die aus der Konvexitätsannahme resultierende Erweiterung der Technologiemenge ist dunkelgrau hinterlegt. Dabei muss allerdings darauf hingewiesen werden, dass diese dunkelgrau eingefärbten Bereiche nicht zwangsläufig in der Realität und auf Basis der vorliegenden Technologie erreichbar sind. Es handelt sich um eine mathematische Annahme.

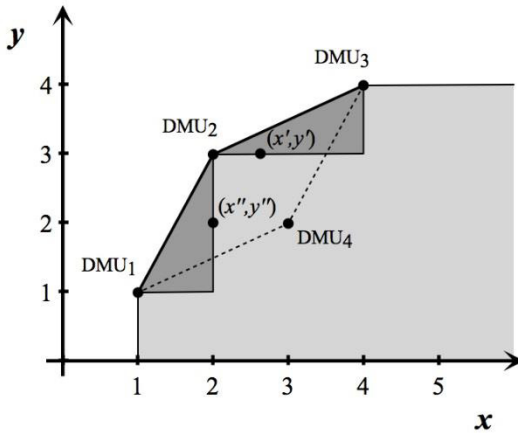


Abbildung 2.6: Technologiemenge bei Verschwendbarkeit und Konvexität  
(Quelle: eigene Darstellung)

Der effiziente Rand eines Technologieraums stellt die Menge aller (auf Basis der Annahmen möglichen) effizienten Aktivitäten dar und entspricht somit der Produktionsfunktion bzw. hier dem approximativ ermittelten Verlauf der Produktionsfunktion. Dieser (approximierte) effiziente Rand der Technologiemenge ist in Abbildung 2.7 gestrichelt dargestellt. Eine Aktivität ist dabei genau dann effizient, wenn es nicht möglich ist, einen Input bei gegebenem Output zu reduzieren oder einen Output bei gegebenem Input zu erhöhen. Wir belassen es hier bei der rein verbalen Erläuterung und werden an geeigneter Stelle noch einmal auf diese Fragestellung zurückkommen.

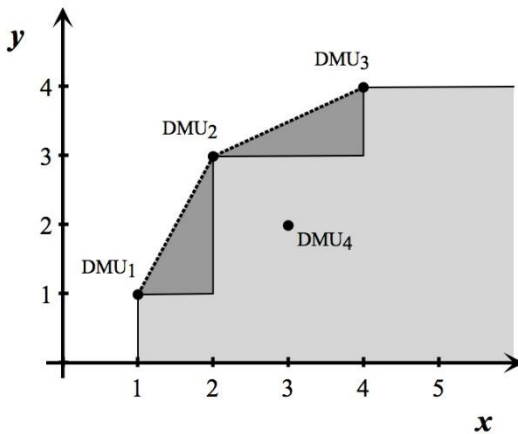


Abbildung 2.7: Der effiziente Rand der Technologiemenge  
(Quelle: eigene Darstellung)

Zur Veranschaulichung des approximativen Charakters des effizienten Randes enthält Abbildung 2.8 zusätzlich eine ertragsgesetzliche Produktionsfunktion mit dem typischen S-förmigen Verlauf (siehe auch Abbildung 2.3 auf Seite 16). In diesem Beispiel zeigt sich, dass der effiziente Rand des Technologieraums zwischen den DMUs 1 und 2 zu optimistisch eingeschätzt wurde, es wurden also Aktivitäten als erreichbar eingestuft, die bei der vorliegenden Technologie nicht realisierbar sind – dieser Bereich ist im Graphen blau eingefärbt. Zwischen den DMUs 2 und 3 findet sich ein (rot eingefärbter) Bereich, der eine partiell zu pessimistische Einschätzung der Produktionsfunktion deutlich macht, d. h. die DMUs werden in diesem Bereich als zu effizient eingeschätzt, die DEA „legt die Latte zu tief“. Faktisch lassen sich diese Effekte allerdings nur bei bekannter Produktionsfunktion feststellen – da in dieser Arbeit von einer unbekannten Produktionsfunktion ausgegangen wird, ziehen wir die Konvexitätsannahme als probates Mittel zur Approximation der Produktionsfunktion heran.

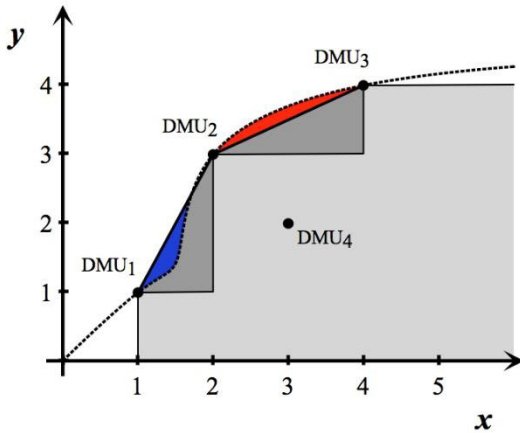


Abbildung 2.8: Konvexe Technologie im Kontrast zu einer ertragsgesetzlichen Produktionsfunktion  
(Quelle: in Anlehnung an [SCH2000, S. S. 48])

#### (4) Minimale Hülle/Extrapolation

$T$  ist die kleinste Technologie, die den Axiomen (1) bis (3) genügt.

In diesem Axiom wird schon ein Bezug zum Begriff der in Unterkapitel 2.2 zu behandelnden Data Envelopment Analysis deutlich: Die Technologie  $T$  umhüllt (englisch: „to envelop“) die unter den genannten Axiomen und auf Basis der beobachteten Aktivitäten der DMUs möglichen In- und Outputkombinationen. Diese Technologie ist nur genau so groß, wie sie sein muss, um die Beobachtungen einzuhüllen und gleichzeitig den genannten Axiomen zu genügen. Dies wirkt sich auf die Effizienzwerte der DMUs dahingehend aus, dass bei einer kleineren Technologiemenge auch der effiziente Rand tendenziell näher an die Aktivitäten der DMUs rückt, was zu steigenden Effizienzwerten führt. Dieser Effekt ist durchaus erwünscht, da die DMUs an erreichbaren Zielen gemessen werden sollen. Allerdings lassen sich Überschreitungen der realen Produktionsfunktion durch die approximierte Technologiemenge auch unter Berücksichtigung von Axiom (4) nicht ausschließen (siehe dazu wiederum Abbildung 2.8).

Auf Basis der angeführten Axiome ergibt sich für die Technologie die folgende mathematische Repräsentation:

$$T = \left\{ (\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \sum_{j=1}^J \lambda_j \mathbf{x}_j, \mathbf{y} \leq \sum_{j=1}^J \lambda_j \mathbf{y}_j, \sum_{j=1}^J \lambda_j = 1, \lambda_j \geq 0 \forall j \right\}$$

### (5) *Radiale Erweiterung*

In den vorangegangenen Ausführungen ist anschaulich geworden, dass es sich bei den Axiomen um Bildungsgesetze der Technologiemenge handelt. Diese werden nun noch um einen Aspekt ergänzt; die Radiale Erweiterung kennt vier verschiedene Ausprägungen:

- (a) Radiale Unbeschränktheit: Wenn  $(\mathbf{x}, \mathbf{y}) \in T$ ,  
dann ist auch  $\alpha(\mathbf{x}, \mathbf{y}) \in T \forall \alpha > 0$ ,
- (b) Radiale Extrapolierbarkeit: Wenn  $(\mathbf{x}, \mathbf{y}) \in T$ ,  
dann ist auch  $\alpha(\mathbf{x}, \mathbf{y}) \in T \forall \alpha \geq 1$ ,
- (c) Radiale Reduktion: Wenn  $(\mathbf{x}, \mathbf{y}) \in T$ ,  
dann ist auch  $\alpha(\mathbf{x}, \mathbf{y}) \in T \forall \alpha \in [0,1]$ ,
- (d) Radiale Beschränktheit: Keine der Aussagen (a) bis (c) ist gültig.

Abbildung 2.9 stellt die graphischen Auswirkungen der Varianten (a) bis (d) mithilfe je eines Graphen pro Variante dar. Diese Graphen unterstellen die Geltung der zuvor festgelegten Axiome (1) bis (4).



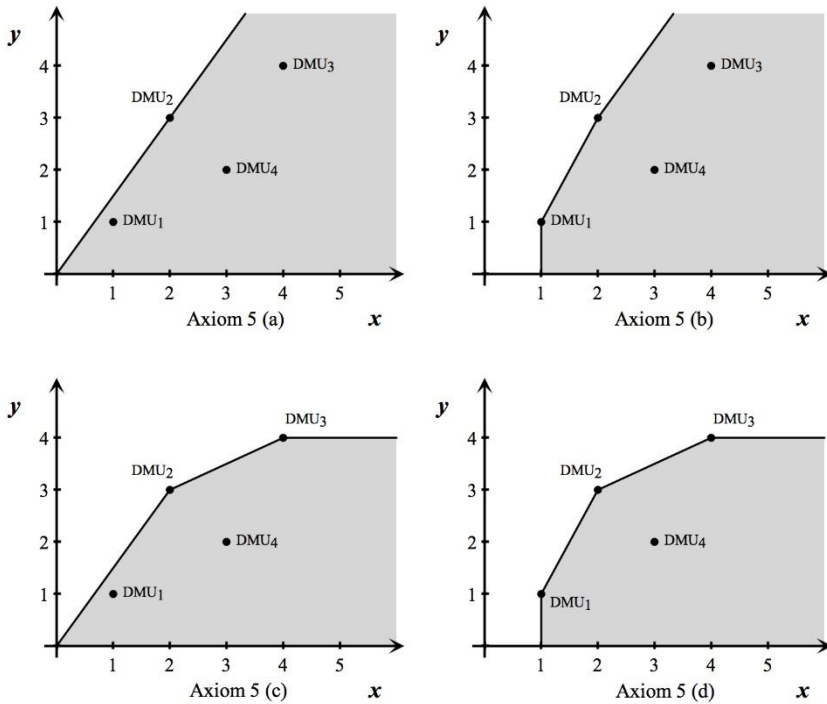


Abbildung 2.9: Technologien bei Variation des Axioms „Radiale Erweiterung“  
(Quelle: in Anlehnung an [ROD2011, S. 8])

An dieser Stelle sei nachträglich darauf hingewiesen, dass die vorangegangenen Abbildungen 2.4 bis 2.8 auf der Annahme von  $\alpha = 1$ , also dem Axiom 5 (d), basieren. In der Literatur wird im Zusammenhang mit der Wahl des  $\alpha$  von der sogenannten Skalenertragslage gesprochen. Dieses Thema wird im weiteren Verlauf dieser Arbeit noch mehrmals von Bedeutung sein und dabei genauer erläutert werden.

Unter Berücksichtigung von Axiom (5) ergibt sich als mathematische Repräsentation der Technologie:

$$T = \left\{ (x, y) \mid x \geq \alpha \sum_{j=1}^J \lambda_j x_j, y \leq \alpha \sum_{j=1}^J \lambda_j y_j, \text{ für alle } \lambda_j \text{ und } \alpha \text{ mit } \sum_{j=1}^J \lambda_j = 1, \lambda_j \geq 0 \forall j, \alpha \text{ gemäß d. rad. Erw. (a), (b), (c) oder (d)} \right\}$$

### 2.1.2.2 Effizienzmessung und Effizienzmaße

Die vergleichende Beurteilung von ökonomischen Einheiten (DMUs) auf Basis von Effizienzmaßen wird als **Effizienzanalyse** bezeichnet [CAH1998, S. 228]. Im Normalfall wird eine Effizienzanalyse mit der Absicht durchgeführt, Ineffizienzen zu erkennen und die Effizienz auf Basis der gewonnenen Erkenntnisse zu steigern oder aber die Wirkung vergangener Effizienzsteigerungsmaßnahmen zu überprüfen. Natürlich muss eine erfolgreiche und fundierte Effizienzanalyse eine Reihe von Schritten durchlaufen und müssen in jedem einzelnen dieser Schritte Annahmen und Festlegungen getroffen werden, die für die Untersuchung und deren Ergebnisse von entscheidender Bedeutung sind. Genau das ist in Bezug auf einen speziellen Bereich der Effizienzanalyse, die Data Envelopment Analysis, ein zentrales Thema dieser Arbeit. In diesem Abschnitt sollen jedoch zuerst einige etablierte und historisch relevante Methoden der Effizienzmessung vorgestellt und in den Gesamtzusammenhang eingeordnet werden.

Immer dann, wenn die Erwartungen an eine Effizienzanalyse über binäre Beurteilungen der DMUs als „effizient“ oder „nicht effizient“ hinausgehen und eine Quantifizierung der eventuellen Ineffizienzen gefordert ist, bedarf es geeigneter Effizienzmaße [SCH2000, S. 75]. Im Rahmen solcher Effizienzmaße kommt der im vorangegangenen Abschnitt eingeführten Technologiemenge eine besondere Bedeutung zu: Wie bereits erläutert, handelt es sich bei den hier betrachteten Effizienzmaßen um relative Größen, es wird also ein Bezug zwischen den Aktivitäten der DMUs und entsprechenden Vergleichsgrößen als Bezugspunkt hergestellt. Diese Vergleichsgrößen stellen immer die unter der gegebenen Technologie bestmöglichen Input-/Output-Kombinationen dar und werden Best-practice-Einheiten genannt. Unabhängig davon, ob die Produktionsfunktion gegeben oder

unbekannt ist, muss sich eine Best-practice-Einheit auf dem effizienten Rand befinden und gilt selbst als effizient.

Zur Messung der Ineffizienzen einzelner DMUs gilt es, den genauen Abstand der jeweiligen DMU zum effizienten Rand der Technologiemenge zu bestimmen. Bereits im Jahr 1951 führte **DEBREU** einen vektorbasierten Ansatz eines Effizienzmaßes, des „Coefficient of Resource“, ein, nach dem die Ineffizienz als Abstand zwischen der Aktivität der betreffenden DMU und einer effizienten Bezugsaktivität definiert ist [DEB1951]. Abbildung 2.10 veranschaulicht diese vektorbasierte Darstellung der Ineffizienz an den bereits bekannten vier Beispiel-DMUs. Dabei wird konkret die Ineffizienz der DMU<sub>4</sub> auf Basis einer inputorientierten Betrachtung graphisch dargestellt. Auf die In- und Outputorientierung unterschiedlicher Data Envelopment Analysis-Modelle sowie die Entscheidungsgrundlagen und -konsequenzen dieses Modellkriteriums wird im weiteren Verlauf dieser Arbeit noch mehrfach eingegangen werden. An dieser Stelle soll lediglich festgehalten werden, dass die graphische Darstellung von einer inputorientierten Betrachtung ausgeht, also davon, dass z. B. die DMU<sub>4</sub> die erzielbaren Effizienzverbesserungen durch Reduktion der eingesetzten Inputs erreichen wird.

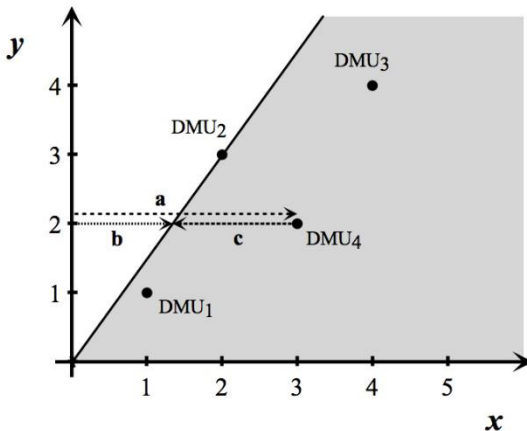


Abbildung 2.10: Graphische Herleitung einer Ineffizienz  
(Quelle: eigene Darstellung)

Die approximative Produktionsfunktion nach dem Modell der radialen Unbeschränktheit stellt dabei den effizienten Rand dar – alle auf dieser Funktion liegenden DMUs sind somit effizient. Die Technologiemenge beschränkt sich also auf den Raum auf und unterhalb der Produktionsfunktion. Der Leser erkennt sofort, dass in der Graphik das Axiom 5 (a) (siehe Abbildung 2.9 auf Seite 25), also  $\alpha > 0$ , unterstellt wurde. DMU<sub>4</sub> liegt deutlich sichtbar unterhalb der Produktionsfunktion und ist somit zwangsläufig ineffizient. Um diese Ineffizienz graphisch herzuleiten wird nun die Frage gestellt, wohin sich die Aktivität dieser DMU verlagern müsste, um effizient zu sein. Im hier unterstellten Fall der Inputorientierung müsste also Input reduziert werden bis die Aktivität der DMU<sub>4</sub> auf der Produktionsfunktion liegt – dies entspricht in Abbildung 2.10 dem Vektor **c**.

Zur Bestimmung der Ineffizienz und damit auch des Effizienzwertes gilt es zuerst den Vektor **a** zu bestimmen, der dem gesamten von der DMU<sub>4</sub> eingesetzten Input entspricht. Dieser Vektor **a**, abzüglich der Inputmenge, die auch von der effizienten Best-practice-Einheit benötigt würde (Vektor **b**) ergibt die einzusparende Inputmenge (Vektor **c**). Dieser Vektor **c** stellt also einen absoluten Wert der Ineffizienz dar. Setzt man seine Länge ins Verhältnis zur Länge von Vektor **a**, ergibt sich ein Ineffizienzwert zwischen 0 und 1:

$$\text{Ineffizienz} = \frac{|\mathbf{a} - \mathbf{b}|}{|\mathbf{a}|}$$

Da die vollständige Effizienz einer Einheit über einen Effizienzwert von 1 ausgedrückt wird, ergibt sich folgendes Effizienzmaß:

$$\text{Effizienz} = 1 - \frac{|\mathbf{a} - \mathbf{b}|}{|\mathbf{a}|} = \frac{|\mathbf{b}|}{|\mathbf{a}|}$$

Diese Vorgehensweise zur Bestimmung von Ineffizienzen und Effizienzen verschiedener DMUs kann nicht nur bei bekannter Produktionsfunktion Anwendung finden. Zwar ist die Produktionsfunktion zur Bestimmung der individuellen Best-practice-Vergleichseinheit(en) einer betrachteten DMU unerlässlich, doch ist es dabei unerheblich, ob es sich um eine tatsächliche Produktionsfunktion oder aber um eine Approximation derselben handelt.

**FARRELL** führte 1957 auf Basis dieser vektorbasierten Methodik ein Verfahren zur Bestimmung der (produktiven) Effizienz bei radialer Unbeschränktheit ein [FAR1957]. Auch er bezog sich auf den effizienten Rand der gegebenen Technologie: "it is far better to compare performances with the best actually achieved than with some unattainable ideal" [FAR1957, S. 255]. Hier spricht FARRELL von realen Best-practice-Einheiten als Vergleichsmaßstab für die einzelnen DMUs, statt fiktiver Idealeinheiten (Anmerkung des Autors: außerhalb der Technologie). Bei dem Effizienzmaß nach FARRELL handelt es sich also um eine relative Effizienz mit realen Best-practice-Einheiten als effizientem Bezugspunkt.

Weiterhin bietet das Effizienzmaß nach FARRELL die Möglichkeit, Sachverhalte mit mehreren Inputs und Outputs abzubilden. Dazu werden die einzelnen In- und Outputs mit Bedeutungs-gewichten  $\mathbf{v}_j^T = (v_{j1}, \dots, v_{jm}, \dots, v_{jM}) \in \mathbb{R}_+^M$  auf Inputseite und  $\mathbf{u}_j^T = (u_{j1}, \dots, u_{js}, \dots, u_{jS}) \in \mathbb{R}_+^S$  auf Outputseite versehen:

$$\text{Effizienz einer DMU}_j = \frac{u_{j1}y_{j1} + u_{j2}y_{j2} + \dots}{v_{j1}x_{j1} + v_{j2}x_{j2} + \dots}$$

Allerdings gelten hier dieselben Bedeutungsgewichte für alle DMUs, die über durchaus unterschiedliche Stärken und Schwächen verfügen. Insofern liegt gerade in dieser Festlegung eines für alle DMUs gleichermaßen geltenden Gewichtungssystems ein Problem. Im weiteren Verlauf dieser Arbeit wird dieses Thema noch einmal aufgegriffen und eine Alternative aufgezeigt, die der Unterschiedlichkeit der DMUs besser Rechnung trägt.

Eine weitere Besonderheit der Effizienzmessung nach FARRELL lag bei ihrer Einführung darin, dass er die Effizienz als Produkt aus der *technischen Effizienz* (TE) und der *allokativen Effizienz* (AE) beschreibt (Die vektorbasierte Methode zur Ermittlung der technischen Effizienz wurde in den vorangegangenen Ausführungen zur Effizienz nach DEBREU bereits dargestellt.). Um die Fragestellung der allokativen Effizienz graphisch veranschaulichen zu können, unterstellen wir einen Transformationsprozess mit zwei Inputs und einem Output. Die Darstellung erfolgt zweidimensional mit den Achsen  $x_1$  und  $x_2$ , bei Normierung des Outputs  $y$  auf den Wert 1. Des Weiteren wird unterstellt, die (für alle DMUs einheitlichen) Faktorpreise der Inputs seien bekannt, so dass eine belastbare Kostenfunktion als Vergleichsmaßstab einzelner Aktivitäten herangezogen werden kann. Es sei darauf hingewiesen, dass die hier angeführten DMUs nicht identisch sind mit den vorher verwendeten Beispiel-DMUs:

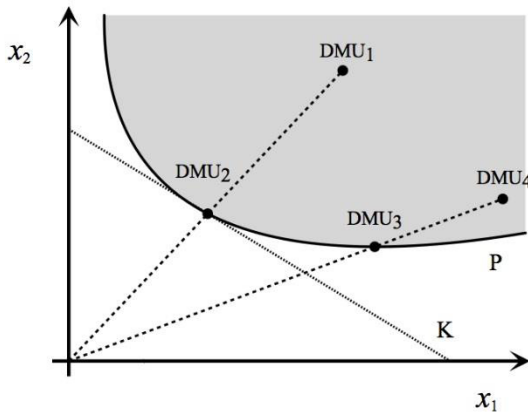


Abbildung 2.11: Technische und allokativen Effizienz  
(Quelle: in Anlehnung an [FAR1957])

Da in dieser graphischen Darstellung (Abbildung 2.11) der Output auf den Wert 1 normiert wurde, entsprechen die eingezeichneten Aktivitäten aller DMUs dem-

selben Outputniveau, sie unterscheiden sich lediglich hinsichtlich der Inputmengen  $x_1$  und  $x_2$ , die auf die Erstellung des Outputs verwendet werden. Es handelt sich also um eine rein inputorientierte Betrachtung. Bei der Isoquante P handelt es sich um die Produktionsfunktion. Alle auf dieser Produktionsfunktion liegenden Aktivitäten sind technisch effizient im Sinne von DEBREU [DEB1951] und FARRELL [FAR1957]. Der Technologieraum liegt in der hier vorliegenden Darstellung oberhalb der Isoquante P, die Isoquante selbst bildet den effizienten Rand dieses Technologieraums. Vom Vorliegen einer allokativen Effizienz kann genau dann gesprochen werden, wenn eine Aktivität nicht nur technisch effizient ist, sich also auf dem effizienten Rand der Produktionsfunktion befindet, sondern gleichzeitig auch das Verhältnis der Faktoreinsatzmengen der Inputs vor dem Hintergrund derer Preise optimal wählt. Die Strecke K bildet die Isokostengerade, abgeleitet als Tangente an der Aktivität der DMU<sub>2</sub> auf der Produktionsfunktion P. Diese Isokostengerade drückt das Verhältnis der Preise der beiden Inputfaktoren  $x_1$  und  $x_2$  zueinander aus. Demzufolge sind in Abbildung 2.11 sowohl DMU<sub>2</sub> als auch DMU<sub>3</sub> technisch effizient, doch lediglich DMU<sub>2</sub> ist gleichzeitig auch allokativ effizient. Die DMUs 1 und 4 sind beide technisch ineffizient aber DMU<sub>1</sub> agiert dennoch allokativ effizient. Das lässt sich dadurch erklären, dass sie auf demselben Ursprungsstrahl liegt wie DMU<sub>2</sub> und somit auch die gleichen Faktorelationen einsetzt.

Abbildung 2.12 zeigt am Beispiel von DMU<sub>4</sub> die technische und die allokativen Ineffizienz graphisch auf:

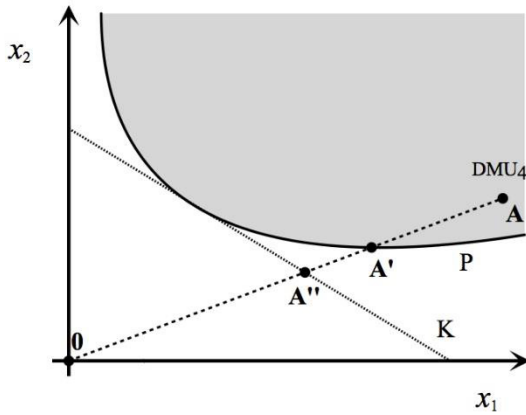


Abbildung 2.12: Beispielhafte Zerlegung der Effizienz  
(Quelle: eigene Darstellung)

DMU<sub>4</sub> ist technisch und allokativ ineffizient. Dabei stellt das Streckenverhältnis  $A$  bis  $A'$  die technische Ineffizienz dar, während sich die technische Effizienz konkret als

$$TE = \frac{|0A'|}{|0A|}$$

darstellt. Das Streckenverhältnis  $A'$  bis  $A''$  bildet die alloкатive Ineffizienz ab, dementsprechend ist die alloкатive Effizienz:

$$AE = \frac{|0A''|}{|0A'|}$$

Die ökonomische Effizienz  $OE$  ist allgemein definiert als:

$$OE = TE \times AE$$

In Bezug auf Abbildung 2.12 bedeutet das, dass  $OE$  mit folgender Gleichung darstellbar ist:

$$OE = \frac{|0A'|}{|0A|} \frac{|0A''|}{|0A'|} = \frac{|0A''|}{|0A|}$$



FARRELS Effizienzbetrachtung wurde 1985 von FÄRE ET AL. [FGK1985] um den Aspekt der Skaleneffizienz erweitert. Mit der Frage nach der Skaleneffizienz wird die generelle Annahme aufgegeben, dass eine DMU unabhängig von Ihrer Größe dieselbe Produktivität erreichen kann und insofern die Größe der Einheit für die erreichbare Produktivität nicht relevant ist. In der Realität hat sich gezeigt, dass in verschiedenen Branchen und Tätigkeitsfeldern die Größe der betrachteten Einheiten durchaus einen Einfluss auf die Effizienz bzw. auf die erzielbare Produktivität hat. Dies sei veranschaulicht am Beispiel einer ertragsgesetzlichen Produktionsfunktion:

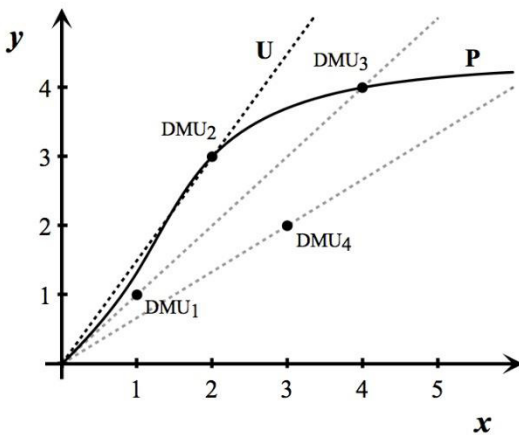


Abbildung 2.13: Skaleneffizienz  
(Quelle: eigene Darstellung)

In Abbildung 2.13 wird der zweidimensionale Fall mit einem In- und einem Output sowie den bekannten DMUs 1 bis 4 dargestellt. Leicht lässt sich erkennen, dass DMU<sub>2</sub> mit der höchsten Produktivität arbeitet, da der Ursprungsstrahl  $U$  durch ihre Aktivität die höchste Steigung aufweist. Bei Unterstellung konstanter

Skalenerträge, also der Annahme, dass eine DMU immer die gleiche Produktivität erreichen kann, sowohl bei Schrumpfung als auch Expansion der Tätigkeit, müssten alle Punkte auf diesem Ursprungsstrahl für die DMU<sub>2</sub> erreichbar sein. Die Produktionsfunktion P zeigt jedoch, dass im vorliegenden Fall einer ertragsgesetzlichen Produktion die erreichbare Produktivität durchaus vom Niveau der In- und Outputs abhängig ist und somit von sogenannten variablen Skalenerträgen ausgegangen werden muss (zur ertragsgesetzlichen Produktionsfunktion siehe z. B. [GUT1983, S. 303 ff.]). So liegen die DMUs 2 und 3 trotz unterschiedlicher Produktivitäten gleichermaßen auf der Produktionsfunktion und somit auf dem effizienten Rand der Technologiemenge – sind also beide effizient obwohl ihre Produktivitäten unterschiedlich sind. Dementsprechend kann DMU<sub>3</sub> ihre Produktivität nur durch Anpassung ihrer Größe steigern, konkret gesagt, durch Schrumpfung. DMU<sub>1</sub> hingegen kann nur durch Expansion die maximale Produktivität erreichen.

Vor diesem Hintergrund muss der Begriff der technischen Effizienz in Bezug auf die Skalenertragslage weiter ausdifferenziert werden: Im Rahmen konstanter Skalenerträge spricht man von „globaler“ technischer Effizienz (GTE) einer DMU, während man bei Unterstellung variabler Skalenerträge von „lokaler“ oder „purer“ technischer Effizienz (LTE) spricht. Die Skaleneffizienz ergibt sich über den Quotienten der beiden technischen Effizienzen:

$$SE = \frac{GTE}{LTE}$$

Zur Ermittlung der Skaleneffizienz müssen also sowohl die lokale als auch die globale technische Effizienz einer DMU ermittelt werden. Ergibt sich für die Skaleneffizienz ein Wert von 1, so arbeitet die betreffende DMU bereits größenoptimal.

Abbildung 2.14 zeigt die DMUs 2 und 5 in Verbindung mit der bekannten Produktionsfunktion:

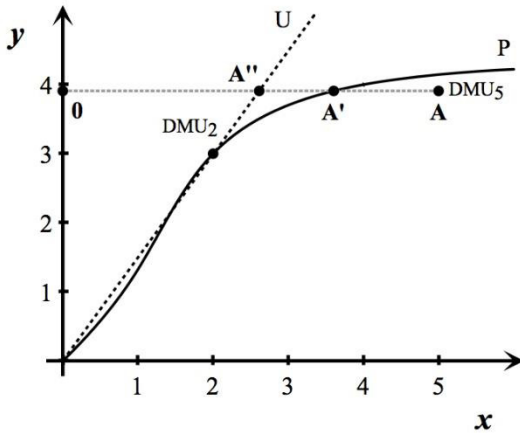


Abbildung 2.14: Graphische Darstellung von technischer und Skaleneffizienz  
(Quelle: eigene Darstellung)

Konkret stellen sich die genannten Effizienzbegriffe am graphischen Beispiel der DMU<sub>5</sub> wie folgt dar:

$$GTE = \frac{|OA''|}{|OA|}$$

$$LTE = \frac{|OA'|}{|OA|}$$

$$SE = \frac{|OA''|}{|OA|} \frac{|OA|}{|OA'|} = \frac{|OA''|}{|OA'|}$$

DMU<sub>2</sub> ist sowohl lokal technisch als auch global technisch effizient, folglich ist sie skaleneffizient. Am Beispiel von DMU<sub>5</sub> lässt sich erkennen, dass die bislang als technische Ineffizienz bezeichnete Distanz (A bis A'') nun unterteilt ist in die lokale technische Ineffizienz (A bis A') und die Skaleneffizienz (A' bis A'') – die technische Ineffizienz wurde also in zwei Komponenten aufgespalten.

Bei gleichzeitiger Betrachtung der allokativen Fragestellung lässt sich die ökonomische Effizienz definieren als:

$$OE = AE \times LTE \times SE$$

Formal ist eine Aktivität  $(x, y)$  genau dann inputorientiert effizient nach DEBREU und FARRELL, wenn

$$(x, y) \in T \text{ und } \forall h < 1 \text{ gilt } (hx, y) \notin T.$$

Analog ist eine Aktivität outputorientiert effizient, wenn gilt:

$$(x, y) \in T \text{ und } \forall g > 1 \text{ gilt } (x, gy) \notin T$$

Die technische Effizienz auf Basis von DEBREU und FARRELL kann als eine Vorstufe oder Basis der technischen Effizienz nach KOOPMANS betrachtet werden: Nach DEBREU und FARRELL sind alle DMUs, die sich auf dem Rand der Technologie befinden, auch als effizient zu bezeichnen. Zur Verdeutlichung der Effizienzbegriffe nach DEBREU/FARRELL und PARETO/KOOPMANS zeigt Abbildung 2.15 im Vergleich zu den Abbildungen 2.11 und 2.12 eine veränderte approximative Produktionsfunktion: DMU<sub>1</sub> ist offensichtlich ineffizient und könnte sich durch radiale Inputreduktion auf den Punkt B' bewegen. DMU<sub>4</sub> ist ebenfalls ineffizient und könnte analog auf den Punkt A' und somit auf den Rand der Technologie gelangen. Damit wäre sie technisch effizient im Sinne von DEBREU und FARRELL.

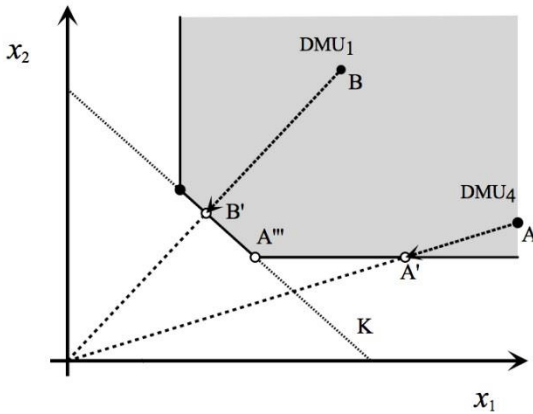


Abbildung 2.15: Effizienz und „schwache“ Effizienz  
(Quelle: eigene Darstellung)

Abbildung 2.15 zeigt aber auch, dass  $DMU_4$  (im Gegensatz zu  $DMU_1$ ) nach der radialen Inputreduktion ihren Input  $x_1$  weiter reduzieren könnte, ohne dadurch Output zu verlieren; sie könnte sich also auf dem Rand der Technologie von  $A'$  bis zum Punkt  $A'''$  bewegen – man spricht in einem solchen Fall von partieller Ineffizienz. Im Rahmen der noch einzuführenden Data Envelopment Analysis wird eine solche Effizienz, wie sie in Punkt  $A'$  vorliegt, als „schwache“ Effizienz bezeichnet.

Die technische Effizienz nach KOOPMANS schließt die genannten schwachen Effizienzen aus; sie stand bereits für die einführende Umschreibung des Effizienzbegriffes in Abschnitt 2.1.1 Pate und soll Grundlage der weiteren Ausführungen sein: Nach KOOPMANS gilt eine Einheit erst dann als technisch effizient, wenn sie nicht nur auf dem Rand des Technologieraums liegt, sondern gleichzeitig keinen Schlupf in der Produktion aufweist, d. h. wenn es dieser DMU nicht möglich wäre, bei gleichen Inputs mehr Output zu erwirtschaften oder bei gleichen Outputs den Input zu reduzieren [KOO1951, S. 60].

KOOPMANS' Überlegungen basieren auf dem bereits 1897 von PARETO entwickelten „schwachen Wohlfahrtsprinzip“ [PAR1897], weshalb auch häufig von der

PARETO/KOOPMANS-Effizienz gesprochen wird. Mathematisch ausgedrückt ist nach **PARETO/KOOPMANS** eine Aktivität  $(x, y) \in T$  dann als technisch effizient zu bezeichnen, wenn

$$\forall (x', y') \in T \text{ mit } x' \leq x \text{ und } y' \geq y \text{ gilt } (x', y') = (x, y).$$

Insofern ist die technische Effizienz nach DEBREU und FARRELL eine notwendige, nicht jedoch eine hinreichende Bedingung für die Effizienz nach KOOPMANS [KUL2000, S. 44]. Auf die in diesem Zusammenhang relevanten „Schlupfvariablen“ wird der Autor im Rahmen der noch zu behandelnden Data Envelopment Analysis näher eingehen.

Zur weiteren Erläuterung des PARETO/KOOPMANS-Kriteriums sind die Fälle konstanter und variabler Skalenerträge separat zu betrachten. Bei Unterstellung konstanter Skalenerträge, also der Annahme, dass die Größe der DMU keinen Einfluss auf die Produktivität hat, lässt sich für jede DMU eine individuelle Produktionsfunktion (Actual-practice-Produktionsfunktion) darstellen, nämlich den Ursprungsstrahl, der durch die Aktivität der betreffenden DMU verläuft. Alle auf diesem Strahl liegenden Produktionspunkte sind für die DMU bei gleichbleibender Effizienz bzw. Ineffizienz erreichbar. In Abbildung 2.16 sind die DMUs 2 und 3 sowie deren Actual-practice-Produktionsfunktionen  $P_2$  und  $P_3$  dargestellt.

Über diese Produktionsfunktionen lässt sich nun eine vergleichende Aussage über die Effizienzen der DMUs 2 und 3 treffen: Durch Schrumpfung der Produktion von  $DMU_3$  wandert der Produktionspunkt auf  $P_3$  in Richtung Ursprung, bis er zwischen  $b'$  und  $b$  liegt, ohne dass sich dabei die Produktivität von  $DMU_3$  verändert. Im Bereich zwischen  $b'$  und  $b$  zeigt sich, dass  $DMU_3$  eindeutig von  $DMU_2$  dominiert wird, da sie im direkten Vergleich mit mehr Input weniger Output erzielt. Analog ließe sich dasselbe Ergebnis durch Expansion der Aktivität von  $DMU_2$  auf  $P_2$  bis in den Bereich oberhalb von  $c$  ermitteln, wobei die obere Grenze  $c'$  außerhalb des im Graphen dargestellten Bereiches liegt. Abschließend zeigt sich, dass die Actual-practice-Produktionsfunktion mit der größten Steigung die Best-practice-Produktionsfunktion ist [CKH2007, S. 154].

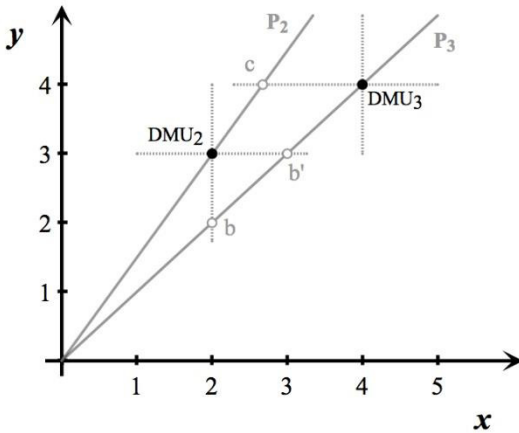


Abbildung 2.16: PARETO/KOOPMANS-Kriterium bei Annahme konstanter Skalenerträge  
(Quelle: in Anlehnung an [CKH2007, S. 153])

Geht man nun jedoch von einem Einfluss der Größe einer DMU auf die Produktivität aus, muss die Betrachtung auf der Grundlage variabler Skalenerträge erfolgen; dementsprechend wird die Annahme der radialen Unbeschränktheit gemäß Axiom 5 (a) aufgegeben, zugunsten einer radialen Beschränktheit gemäß Axiom 5 (d) (siehe Abbildung 2.9 auf Seite 25). Eine DMU stellt sich nun also dem Vergleich mit einer Technologie, die den Axiomen 1 bis 4 genügt. Um die DMUs unter dieser Prämisse dennoch vergleichen zu können, werden die Konvexkombinationen der DMUs in die Betrachtung mit einbezogen. Zur Identifikation der nach dem PARETO/KOOPMANS-Kriterium effizienten oder ineffizienten Einheiten wird die Frage gestellt, welche DMUs von mindestens einer der anderen DMUs oder Konvexkombination(en) der anderen DMUs dominiert werden. Abbildung 2.17 zeigt die Produktionspunkte der betrachteten Einheit und die umrandete Fläche der Konvexkombinationen:

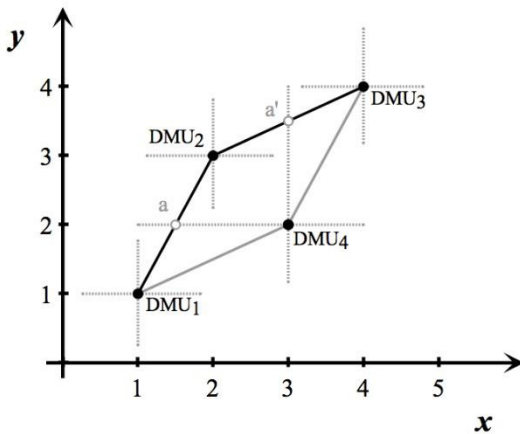


Abbildung 2.17: PARETO/KOOPMANS-Kriterium bei Annahme variabler Skalenerträge  
(Quelle: in Anlehnung an [CKH2007, S. 156])

DMU<sub>4</sub> ist augenscheinlich nicht effizient, da sie von DMU<sub>2</sub> eindeutig dominiert wird, also mehr Input benötigt und gleichzeitig weniger Output erzeugt. Die DMUs 1, 2 und 3 werden von keiner der übrigen DMUs dominiert. Deshalb müssen zur Beurteilung ihrer Effizienz die Konvexkombinationen der DMUs in die Betrachtung einbezogen werden. Eine fiktive Verschiebung der Aktivitäten durch Konvexkombinationen auf dem effizienten Rand des Technologieraums zeigt, dass die DMUs 1, 2 und 3 von keiner anderen DMU dominiert werden und somit nach PARETO/KOOPMANS effizient sind.

Um die Einordnung der verschiedenen Methoden zur Messung relativer Effizienzen bei unbekannter Produktionsfunktion in den Gesamtzusammenhang zu erleichtern, wird im Folgenden auf eine etablierte Form der Kategorisierung zurückgegriffen, wie in Abbildung 2.18 dargestellt.



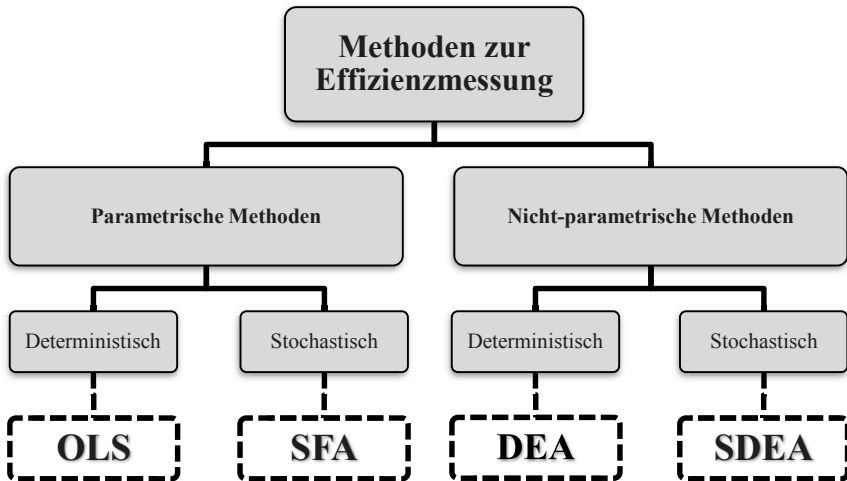


Abbildung 2.18: Methoden der Effizienzmessung  
(Quelle: in Anlehnung an [BUR2008, S. 48])

Parametrische Methoden erfordern einen vorab gegebenen funktionalen Zusammenhang des Transformationsprozesses. Im Allgemeinen wird die dabei unterstellte Funktion mithilfe statistischer Methoden den empirischen Daten angepasst und danach werden die verbleibenden Abweichungen zur tatsächlichen Produktionsfunktion durch statistische Tests ermittelt [BUR2008, S. 48]. Während die deterministischen Ansätze von einer Verlässlichkeit der empirischen Daten ausgehen und somit den gesamten Abstand einer Aktivität vom effizienten Rand als Ineffizienz deuten, beziehen stochastische Verfahren auch Messfehler oder sogenannte „Ausreißer“ in die Betrachtung mit ein [DYS2010, S. 140]. Hier wird der Abstand zum effizienten Rand als eine Mischung aus Ineffizienz und den genannten Fremdeinflüssen interpretiert. Ein Beispiel für parametrische deterministische Effizienzmessung ist die Ordinary-Least-Square-Methode (OLS), die auf Basis der empirischen Daten eine Regressionsgerade ermittelt und diese als effizient definiert. Die Stochastic-Frontier-Analysis (SFA) steht beispielhaft für die parametrischen stochastischen Methoden: Sie bestimmt den effizienten Rand mithilfe eines a priori festgelegten Funktionstyps, der mittels stochastischer Verfahren auf Basis der Datenpunkte ausgerichtet wird und zerlegt die gemessenen

Abstände in Ineffizienzen und Fremdeinflüsse [BUR2008, S. 49 ff.]. Der wichtigste Kritikpunkt an den parametrischen Methoden liegt in der oft willkürlich gewählten Form der Produktionsfunktion [SCG1994, S. 939].

Die nicht-parametrischen Methoden bedürfen keiner vorgelagerten Definition der Produktionsfunktion und kommen erst im Rahmen der Effizienzbestimmung auf Basis der empirischen Daten zu einer Schätzung bzw. Approximation derselben. Die Untergliederung in statistische und deterministische Verfahren erfolgt analog zu den parametrischen Verfahren. Wie in Abbildung 2.18 dargestellt, handelt es sich bei der Data Envelopment Analysis (DEA), die das zentrale Thema dieser Arbeit darstellt, um eine deterministische nicht-parametrische Methode der Effizienzmessung, die in Unterkapitel 2.2 näher erläutert wird. Eine Abwandlung dieser Methode, die Stochastic Data Envelopment Analysis (SDEA), ergänzt die Data Envelopment Analysis um eine Korrektur der Effizienzwerte zur Eliminierung statistisch ermittelter Fehler – insofern wird die SDEA hier als Beispiel für eine stochastische nicht-parametrische Methode genannt. Aufgrund ihrer geringen Bedeutung in der einschlägigen theoretischen und vor allem empirischen Literatur wird die SDEA in der vorliegenden Arbeit nicht näher betrachtet.

Alle genannten Methoden zur Messung relativer Effizienzen sollten gewissen gemeinsamen Anforderungen genügen. Da in der Literatur keine etablierte abgeschlossene Liste solcher Anforderungen existiert, soll an dieser Stelle eine Auswahl der wichtigsten Punkte genügen:

- *Effizienzindikation*

Das Maß der Ineffizienz soll in einem festen Wertebereich liegen, z. B. 0 bis 1, wobei die Grenzen dieses Wertebereichs für Effizienz bzw. vollständige Ineffizienz stehen sollen. Alle Werte dazwischen geben den Grad einer teilweisen Ineffizienz an.

Da außerdem eine einfache und intuitive Interpretierbarkeit des Effizienzmaßes anzustreben ist [SCH2000, S. 80], sollte immer wenn möglich der übliche Wertebereich von 0 (vollständige Ineffizienz) bis 1 (Effizienz) zur Anwendung kommen. Ein solcher allgemeingültiger Wertebereich ist insbesondere auch bei der vergleichenden Interpretation verschiedener Effizienzmaße von Vorteil [CPP1999, S. 17].

- *Relevanz*  
Von entscheidender Bedeutung ist eine möglichst hohe Aussagekraft des Effizienzmaßes bezüglich der vorab festgelegten Fragestellung. Dieser Anspruch ist in allen Phasen der Modellierung zu berücksichtigen und wird sich in seinen verschiedenen Aspekten als roter Faden durch diese Arbeit ziehen.
- *Skaleninvarianz*  
Der Effizienzgrad einer DMU soll unabhängig von den verwendeten – eventuell auch unterschiedlichen – Messeinheiten der In- und Outputs sein.
- *Präferenzkompatibilität bzw. Monotonie*  
Die Effizienzgrade dürfen den bilateralen Dominanzen der DMUs nicht widersprechen. Die dominierende Einheit muss auch den höheren Effizienzwert aufweisen [DYS2010, S. 136 f.]. Dies wird gewährleistet durch den Anspruch, dass der Abstand einer Aktivität vom effizienten Rand bzw. den darauf liegenden Referenz-Aktivitäten einen Indikator für das Ausmaß der Ineffizienz der betreffenden DMU darstellt.
- *Stabilität*  
Der Effizienzwert einer DMU sollte sich durch geringe Veränderungen bzw. Fehler in den empirischen Daten nicht drastisch verändern, sondern stetig sein [SCH2000, S. 86 f.].

Neben der bereits erwähnten Unvollständigkeit muss sich die vorangegangene Aufzählung einen weiteren Kritikpunkt gefallen lassen, nämlich dass die angeführten Anforderungen teilweise inkompatibel miteinander sind [SCH2000, S. 88] [RUS1990, S. 266]. Insofern sollten letztlich für jedes Effizienzmaß individuelle Anforderungen definiert werden, die auch alle Rahmenbedingungen der Untersuchung berücksichtigen. Deshalb wird diese Fragestellung im folgenden Unterkapitel 2.2 bei den Grundlagen der Data Envelopment Analysis noch einmal separat behandelt.

## 2.2 Data Envelopment Analysis

Am Ende des vorangegangenen Abschnitts wurden allgemeine Schwächen und Unzulänglichkeiten von Methoden zur Messung relativer Effizienzen herausgestellt. Nun soll die Data Envelopment Analysis, die im Zentrum der vorliegenden Arbeit steht, von den Grundlagen bis zu gängigen Modellabwandlungen und -erweiterungen vorgestellt werden.

Zum Abschluss dieses Kapitels wird die Data Envelopment Analysis im Rahmen einer kategorisierenden Übersicht in den Gesamtzusammenhang der Effizienzmessmethoden eingeordnet und werden dabei noch einmal die relevanten Eigenschaften und Besonderheiten vergleichend dargestellt.

### 2.2.1 Grundlagen der Data Envelopment Analysis

In Unterkapitel 2.1 wurden Grundlagen und Entwicklung der Effizienzmessung eingehend behandelt. Die Arbeiten von PARETO [PAR1897], KOOPMANS [KOO1951], SHEPHARD [SHE1970], DEBREU [DEB1951] und insbesondere auch FARRELLS Ausführungen zur Messung der Effizienz von Produktionseinheiten [FAR1957] dürfen dabei als ein Fundament für die spätere Entwicklung der Data Envelopment Analysis betrachtet werden bzw. die Data Envelopment Analysis als eine Weiterführung der in diesen Arbeiten eingeführten Anschauungen und Theorien.

Den konkreten Anfang der Data Envelopment Analysis bildet die Dissertation von EDUARDO RHODES aus dem Jahr 1978 [RHO1978], die dann in Zusammenarbeit mit CHARNES und COOPER zum vielbeachteten Artikel im European Journal of Operational Research [CCR1978] führte, bekannt als CCR-Artikel. Seitdem sind zur Data Envelopment Analysis weit über 5.000 Arbeiten veröffentlicht worden, wobei unveröffentlichte Dissertationen, Arbeits- und Diskussionspapiere usw. noch nicht berücksichtigt sind [EPT2008, S. 152].

Bevor jedoch auf dieses CCR-Modell und etablierte Abwandlungen sowie Erweiterungen genauer eingegangen wird, sollen die relevanten Grundlagen und Annahmen der Data Envelopment Analysis angesprochen werden. Die Data Envelopment Analysis wurde bereits in Unterkapitel 2.1 als nicht-parametrisches und deterministisches Verfahren zur Messung von (In)effizienzen vergleichbarer DMUs eingeführt. Sie greift dabei auf In- und Outputs der zu bewertenden DMUs, also *Aktivitäten*, zurück und stellt somit auf realwirtschaftliche Kenngrößen ab. Konkret werden auf Basis dieser Aktivitäten die *Produktionsfunktion* approximativ bestimmt und auf ihr liegende DMUs als effizient definiert. Diese DMUs, auch „Best-practice-Einheiten“ genannt, dienen als Vergleichsmaßstab zur Bestimmung des Effizienzgrades der übrigen DMUs [BGZ1989, S. 275] [SEI2006, S. 3]. Auf eine genaue Beschreibung des *Technologierums*, der in Unterkapitel 2.1 im Rahmen der allgemeinen Vorstellung verschiedener Effizienzmessmethoden eingeführt wurde, soll an dieser Stelle verzichtet werden. Grundsätzlich halten wir jedoch fest, dass der Technologieraum und insbesondere sein effizienter Rand im Rahmen der Data Envelopment Analysis eine übergeordnete Rolle spielen: Der Begriff Data Envelopment Analysis resultiert daraus, dass sich die Technologiemenge als eine Umhüllung aus den beobachteten Aktivitäten der DMUs ableitet [POR2000, S. 111] und ihr effizienter Rand dementsprechend eine „lineare Verbindung von Extremwerten“ darstellt [BUR2008, S. 51] [AND1996]. Ein mittels Data Envelopment Analysis ermittelter Effizienzwert einer DMU sagt aus, wie (in)effizient diese DMU aus Sicht effizienter Vergleichs-DMUs ist. Die beurteilte DMU wird diesem Effizienzwert mit hoher Akzeptanz begegnen, da der angelegte Vergleichsmaßstab real und damit auch erreichbar ist.

Anfänglich wurde die Data Envelopment Analysis in erster Linie zur Effizienzmessung in Non-profit-Sektoren der öffentlichen Hand angewendet. Hier kann sie ihren Vorteil ausspielen, dass sie keiner gegebenen Marktpreise der betrachteten In- und Outputs bedarf. Doch bedeutet dies in keiner Weise eine schlechtere Eignung für andere Anwendungsfälle. Die DEA wurde inzwischen zur Effizienzmessung aller Arten von produktiven Einheiten eingesetzt, wie z. B. Firmen, Abteilungen, Firmenstandorte, Produktionsprozesse oder Teile davon sowie Projekte oder auch makroökonomische Prozesse [DYS2010, S. 141] [ABD2003, S. 91].

Grundsätzlich muss eine Untersuchung mittels Data Envelopment Analysis verschiedenen Anforderungen gerecht werden und gewisse Anwendungsvoraussetzungen erfüllen. Die hier angeführten Punkte gründen auf einer breiten Auswahl von Arbeiten zur Data Envelopment Analysis, insbesondere aber auf DYCKHOFF UND SPENGLER [DYS2010] sowie PETERS [PET2008] und SCHEEL [SCH2000]:

- *Vergleichbarkeit der DMUs*

Über die Notwendigkeit der Vergleichbarkeit der DMUs einer Untersuchung besteht in der Literatur Einigkeit. Auseinander gehen allerdings die Definitionen einer solchen Vergleichbarkeit: Laut DYCKHOFF UND SPENGLER ist die Anforderung relevant, dass alle DMUs durch dieselben In- und Outputs beschrieben werden [DYS2010, S. 141], also in jedem der definierten In- und Outputs vergleichbar sind. Diese Annahme wird unter anderem von SUHREN [SUH1997, S. 9 f.] bestätigt.

Weiterhin müssen die empirischen Daten, die in die Effizienzberechnung mittels der DEA einfließen, immer genau dem Zeitraum zuzuordnen sein, für den eine Effizienzaussage ermittelt werden soll, so dass sich folglich alle Daten der einzelnen DMUs auf denselben Zeitraum beziehen.

Bei diesen beiden Punkten handelt es sich zwar um notwendige, nicht jedoch hinreichende Bedingungen für die Sicherstellung der Vergleichbarkeit aller betrachteten DMUs. DYCKHOFF UND SPENGLER operieren deshalb mit der Bedingung, dass alle betrachteten DMUs mit derselben Technik arbeiten und auf dieser Basis ihre Aktivitäten realisieren [DYS2010, S. 141]. Zwar kommen sie der Erfüllung einer Vergleichbarkeit der DMUs damit schon deutlich näher, doch dürfte diese kryptische Bedingung dem betriebswirtschaftlich orientierten Anwender in der Praxis wenig helfen. Auch deshalb wird der Autor sich dieser Fragestellung noch einmal konkret in Unterkapitel 4.4 widmen.

- *Erwünschte Outputs*

Die Data Envelopment Analysis unterstellt, dass Minderungen der Inputs und Mehrungen der Outputs bei gleichen Rahmenbedingungen zu einer steigenden Effizienz führen. Diese Annahme erfordert zwangsläufig auch, dass

alle Outputs erwünscht sind. Im Fall unerwünschter Outputs müsste eine Reduktion dieses Outputs eine Steigerung der Effizienz zur Folge haben, was dem grundsätzlichen Verständnis der DEA von In- und Outputs zuwiderläuft. Insofern ist bei der Modellgestaltung auch zu berücksichtigen, dass es sich tatsächlich bei allen Outputs um erwünschte Outputs handelt.

Zwar gibt es durchaus Ansätze zur Berücksichtigung unerwünschter Outputs im Rahmen der DEA [SCH2000, S. 148 ff.], doch sollen diese in der vorliegenden Arbeit nur insoweit behandelt werden, wie sie sich innerhalb der klassischen Modellgestaltung berücksichtigen lassen.

- *Wirkungen zwischen In- und Outputs*

Auch dieser Punkt zielt darauf ab, die unterstellten Wirkungszusammenhänge bei allen praktischen Veränderungen von In- bzw. Outputs gewährleisten zu können. Konkret sollte beispielsweise nicht der Gewinn als Outputgröße eines Unternehmens verwendet werden, da sich dieser mit großer Wahrscheinlichkeit auch aus Größen ableitet, die auf der Inputseite bereits eingeflossen sind [SCH2000, S. 17]. Solche Zusammenhänge könnten dazu führen, dass eine Erhöhung eines Inputs beispielsweise die Reduktion eines Outputs zur Folge hätte. Dieses Thema wird in Zusammenhang mit der Auswahl der In- und Outputs in Unterkapitel 4.4 in verschiedenen Aspekten genauer beleuchtet.

- *Unmöglichkeit des Schlaraffenlandes*

Output kann nur unter dem Einsatz von Input erzeugt werden.

- *Beschränktheit der Technologie*

Mit begrenzten Inputmengen lassen sich auch nur begrenzte Outputmengen erzeugen.

Der Vollständigkeit halber bleibt darauf hinzuweisen, dass die genannten Anforderungen an eine Effizienzanalyse mittels Data Envelopment Analysis nicht anstelle, sondern in Ergänzung zu den in Unterkapitel 2.1 erläuterten Axiomen und Grundlagen für Verfahren relativer Effizienzmessung zu betrachten sind.

## 2.2.2 CCR-Modelle mit Input-/Outputorientierung

Nach den grundsätzlichen Ausführungen zur Data Envelopment Analysis soll nun das bereits erwähnte CCR-Modell zur vergleichenden Effizienzmessung von CHARNES, COOPER UND RHODES [CCR1978], das die Grundlage aller späteren Überlegungen zur Data Envelopment Analysis bildet, eingeführt werden. In ihren Ausführungen bauen CHARNES ET AL. auf FARRELLS Ansatz [FAR1957] auf, die Produktionsfunktion mittels einer Einhüllung der beobachteten Aktivitäten zu approximieren [HOF2006, S. 25]. Dabei ermitteln CHARNES ET AL. die Effizienzwerte der einzelnen DMUs durch den Vergleich jeder DMU mit ähnlichen, aber effizienten, Best-practice-DMUs. Diese Best-practice-DMUs determinieren dabei die approximative Produktionsfunktion [BGZ1989, S. 275]. Im Rahmen des CCR-Modells haben die in Abschnitt 2.1.2 definierte Notation sowie die fünf Axiome zur Beschreibung der Technologiemenge volle Gültigkeit. Bezüglich des Axioms 5 unterstellt das CCR-Modell die Variante (a).

Die Anzahl der Dimensionen des Technologieraums entspricht der addierten Anzahl der betrachteten In- und Outputs [SCH1994, S. 154]. Zur anschaulichen Erläuterung der Modellzusammenhänge wird auch weiterhin auf zweidimensionale Beispiele bzw. Darstellungen zurückgegriffen. Abbildung 2.19 stellt die Technologiemenge mit den bekannten vier beispielhaften DMUs mit je einem Input und einem Output im CCR-Modell graphisch dar.

Bezüglich der Festlegung des Modells auf das Axiom 5 (a), also  $\alpha > 0$ , sprechen wir im Rahmen des CCR-Modells von der Annahme konstanter Skalenerträge. Das bedeutet, dass bei einer gleichzeitigen Skalierung der In- und Outputs um einen beliebigen Faktor die Produktivität konstant gehalten werden kann [HKC2002, S. 173]. In Abbildung 2.19 lässt sich das an dem Expansionspfad erkennen, der aus dem Ursprung durch die produktivste DMU, also DMU<sub>2</sub>, verläuft: Dieser Expansionspfad entspricht der unterstellten Produktionsfunktion und bildet folglich den effizienten Rand der Technologie. Er leitet sich aus einer Expansion oder Schrumpfung der Aktivität von DMU<sub>2</sub> bei gleichbleibender Produktivität ab oder, anders gesagt, aus der Annahme, dass es im Rahmen der gegebenen Technologie möglich ist, die Aktivität einer DMU auszuweiten oder



zu schrumpfen, ohne dabei von einer Veränderung der erzielbaren Produktivität ausgehen zu müssen.

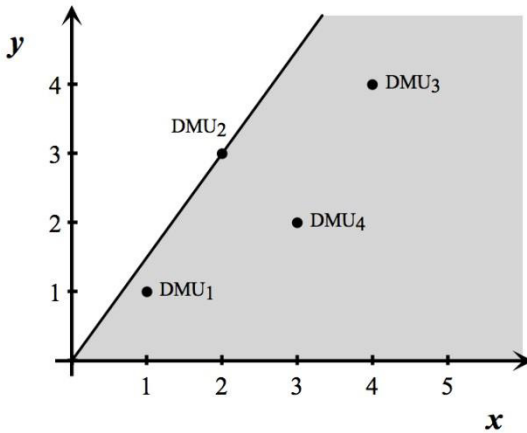


Abbildung 2.19: Technologie im CCR-Modell  
(Quelle: in Anlehnung an [ROD2011, S. 8])

Die Annahme konstanter Skalenerträge ist insbesondere in solchen Branchen realistisch, in denen beispielsweise eine Ausweitung des Geschäfts nicht zu Synergien führt oder, allgemein gesprochen, die Betriebsgröße nicht die Produktivität der Einsatzfaktoren beeinflusst. Diese Gegebenheiten finden sich eher im Dienstleistungssektor als im produzierenden Gewerbe. Da jedoch nur unter den genannten Bedingungen von konstanten Skalenerträgen ausgegangen werden kann, wurde das CCR-Modell zu einem späteren Zeitpunkt weiterentwickelt und an die Annahme variabler Skalenerträge adaptiert. Dieses Modell wird Thema des nächsten Abschnitts sein.

Auf Basis der eingeführten Grundlagen relativer Effizienzmessung und der Data Envelopment Analysis lässt sich Abbildung 2.20 bereits dahingehend interpretieren, dass DMU<sub>2</sub> aufgrund ihrer Lage auf dem effizienten Rand als effizient einzustufen ist. Dementsprechend lässt sich im Graphen erkennen, dass die DMUs

1, 3 und 4 ineffizient sind. Das CCR-Modell liefert die Antwort auf die Frage nach den genauen (relativen) Effizienzwerten der DMUs, wobei es hier direkt zu Anfang zwei Varianten zu unterscheiden gilt. Abbildung 2.20 zeigt diese zwei Varianten am Beispiel von DMU<sub>4</sub> auf:

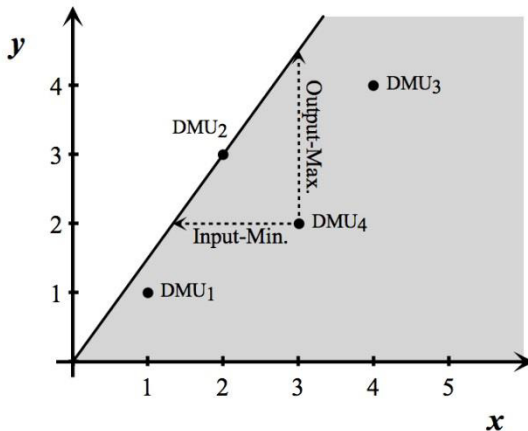


Abbildung 2.20: In- und Outputorientierung im CCR-Modell  
(Quelle: in Anlehnung an [ROD2011, S. 8])

DMU<sub>4</sub> weist eine schlechtere Produktivität als DMU<sub>2</sub> auf und ist damit (relativ) ineffizient. Um (relativ) effizient zu werden, kann sie entweder ihre Outputs bei unveränderten Inputs erhöhen („Output-Max.“ in Abbildung 2.20) oder die Inputs bei unveränderten Outputs reduzieren („Input-Min.“ in Abbildung 2.20), man spricht allgemein von einem **inputorientierten oder outputorientierten Modell**. Im Fall der Inputorientierung wird das Ziel verfolgt, bei konstanten Outputs den Input zu optimieren, also zu reduzieren. Eine ermittelte Effizienz von  $4/9 = 0,\bar{4}$  bedeutet formal ausgedrückt, dass die Inputwerte mit dem Faktor  $0,\bar{4}$  multipliziert werden können, ohne dass dadurch der Output gemindert würde. Wie bereits in Unterkapitel 2.1 erwähnt, handelt es sich bei den im Rahmen der DEA ermittelten Effizienzen um relative Effizienzen, die sich an den jeweiligen Best-practice-DMUs orientieren. Die Aussage, dass die Inputs bei konstanten Outputs

um  $55, \bar{5} \%$  reduziert werden können, leitet sich aus der realen Produktivität der Best-practice-DMU(s) ab. Die Outputorientierung stellt sich analog zur Inputorientierung dar.

Im Rahmen des CCR-Modells ist die Frage nach der Orientierung in mathematischer Hinsicht allerdings von untergeordneter Bedeutung, da sich in beiden Betrachtungen dieselben Effizienzwerte für die einzelnen DMUs ergeben [CKH2007, S. 200]. Die betriebswirtschaftlichen Konsequenzen einer Entscheidung für ein in- oder outputorientiertes Modell werden in Kapitel 4 genauer untersucht. Die formale Darstellung des CCR-Modells soll hier am Beispiel einer inputorientierten Betrachtung erfolgen.

Formal stellt sich die Optimierungsaufgabe des inputorientierten CCR-Modells in der sogenannten **Envelopmentform** für jede DMU<sub>k</sub>,  $k \in \{1, \dots, J\}$ , und auf Basis der in Unterkapitel 2.1 eingeführten Notation folgendermaßen dar:

$$h_k^* = \min h_k \quad (\text{CCR-I-E})$$

$$\text{u. d. N.} \quad h_k \mathbf{x}_k \geq \alpha_k \sum_{j=1}^J \lambda_{kj} \mathbf{x}_j$$

$$\mathbf{y}_k \leq \alpha_k \sum_{j=1}^J \lambda_{kj} \mathbf{y}_j$$

$$\sum_{j=1}^J \lambda_{kj} = 1$$

$$\alpha_k > 0, \lambda_{kj} \geq 0 \quad \forall j$$

$h_k^*$  bildet dabei den Faktor, mit dem jede DMU<sub>k</sub> ihre gesamten Inputs multiplizieren muss, um bei unveränderten Outputs effizient zu werden, also genau auf den Rand der Technologie zu gelangen – anders formuliert:  $h_k^*$  stellt ein Maß für den Abstand der DMU<sub>k</sub> zum Rand dar (die Kennzeichnung von optimalen Lösungen

durch ein hochgestelltes \* wird im weiteren Verlauf dieser Arbeit beibehalten). Im Fall von Beispiel-DMU<sub>4</sub> zeigt der Pfeil „Input-Min.“ in Abbildung 2.20 den entsprechenden Zielpunkt auf dem Rand an. Da in mehrdimensionalen Fällen alle Inputs einer DMU mit demselben Faktor multipliziert werden, spricht man von einer radialen Inputreduktion. Für eine bereits (schwach) effiziente DMU<sub>k</sub> ergibt sich  $h_k^* = 1$ , während die Lösungen für ineffiziente Einheiten zwangsläufig unter 1 liegen, also  $h_k^* < 1$ .

Durch die Variablentransformation  $\mu_{kj} = \alpha_k \lambda_{kj}$  lässt sich die nichtlineare Aufgabe (CCR-I-E) in eine lineare überführen, hier als primales Problem (CCR-I-P) bezeichnet:

$$h_k^* = \min h_k \quad (\text{CCR-I-P})$$

$$\text{u. d. N.} \quad h_k x_k \cong \sum_{j=1}^J \mu_{kj} x_j$$

$$y_k \leq \sum_{j=1}^J \mu_{kj} y_j$$

$$\mu_{kj} \geq 0 \quad \forall j$$

Durch Dualisierung kommt man zur **Multiplifierform** (CCR-I-M):

$$h_k^* = \max U_k^T y_k \quad (\text{CCR-I-M})$$

$$\text{u. d. N.} \quad V_k^T x_k = 1$$

$$U_k^T y_j - V_k^T x_j \leq 0, \quad j = 1, \dots, J$$

$$U_k, V_k \geq 0$$

Die Data Envelopment Analysis ermittelt für jede DMU optimale Gewichtungen der In- und Outputs, so dass die Stärken einer Einheit mehr und die Schwächen weniger ins Gewicht fallen [BGZ1989, S. 275]. Diese individuelle Form der Gewichtung, genannt „Self-appraisal“, führt zu den höchstmöglichen Effizienzwerten der einzelnen DMUs [CCR1978, S. 430] [GUT2005, S. 11] [BDT1991, S. 4] und stellt dadurch die Akzeptanz der Effizienzanalyse bei den einzelnen Teilnehmern sicher [SCH2000, S. 61 f.]. Wenn der Effizienzwert einer DMU gleich 1 ist, dann ist diese DMU bei den für sie optimalen Gewichten im Vergleich zu den anderen DMUs effizient. Bei einem Effizienzwert  $< 1$  ist eine andere DMU selbst unter diesen Gewichten effizienter.

Mit dem Thema vertraute Leser erkennen in der zweiten Nebenbedingung des Optimierungsproblems in Multiplierform (CCR-I-M) bereits den Nenner des Quotientenproblems. Durch eine weitere Variablentransformation mit einem beliebigen  $t_k > 0$  und  $v_k = t_k^{-1} V_k$  sowie  $u_k = t_k^{-1} U_k$  wegen  $v_k^T x_k = t_k^{-1}$  sowie  $u_k^T y_j - v_k^T x_j \leq 0$  (vgl. hierzu [ROR2011, S. 149]) verlagert man die zweite Nebenbedingung in die Zielfunktion und gelangt somit von der Multiplierform zum **Quotientenproblem**:

$$\begin{aligned}
 h_k^* &= \max \frac{\mathbf{u}_k^T \mathbf{y}_k}{\mathbf{v}_k^T \mathbf{x}_k} && \text{(CCR-I-Q)} \\
 \text{u. d. N.} \quad &\frac{\mathbf{u}_k^T \mathbf{y}_j}{\mathbf{v}_k^T \mathbf{x}_j} \leq 1, \quad j = 1, \dots, J \\
 &\mathbf{u}_k, \mathbf{v}_k \geq \mathbf{0}
 \end{aligned}$$

Zwar lassen sich mit dem hier vorgestellten CCR-Modell konkrete Effizienzwerte für die einzelnen DMUs bestimmen und weiterhin lässt sich aus diesen Werten schließen, in welchem Umfang die Inputs radial reduziert werden müssen, damit eine DMU effizient wird. Dennoch besteht nach der entsprechenden Reduktion der Inputs oder, im outputorientierten Modell, nach Ausweitung der Outputs, die Möglichkeit partieller Ineffizienzen, sogenannter Schlüpfе oder **Slacks** [BUR2008, S. 18].

Aufbauend auf den Ausführungen zu den unterschiedlichen Effizienzbegriffen in Unterabschnitt 2.1.2.2, insbesondere Abbildung 2.15, zeigt Abbildung 2.21 die Problematik der Slacks anhand vier beispielhafter DMUs 1 bis 4 mit je zwei Inputs und einem Output. Zugunsten einer Darstellung im zweidimensionalen Raum wurden die Aktivitäten wiederum auf den Output normiert. Durch die radiale Reduktion der Inputs bewegen sich die beiden nicht effizienten DMUs 1 und 4 in Richtung Ursprung, bis sie die Grenze des Technologieraums erreichen. Im Fall von DMU<sub>1</sub> führt das zu Punkt B', der sich auf dem effizienten Rand befindet. DMU<sub>4</sub> hingegen verändert sich zu Punkt A', der zwar auf dem Rand der Technologie, nicht jedoch auf dem effizienten Rand liegt und eindeutig von DMU<sub>3</sub> dominiert wird. Man sieht deutlich den Abstand zur Isokostengeraden K, die die optimale Einsatzrelation der beiden Inputs  $x_1$  und  $x_2$  ausdrückt. DMU<sub>4</sub> kann von Punkt A' aus seinen Input  $x_1$  reduzieren, ohne dabei Output zu verlieren, genau bis er auf Punkt A''' gelangt, auf dem sich auch DMU<sub>3</sub> befindet. Formal ausgedrückt hat der Input  $x_1$  im Punkt A' einen Schattenpreis von 0 bzw. liegt ein Slack  $> 0$  vor. Insofern hat das CCR-Modell für DMU<sub>4</sub> zwar eine optimale Empfehlung für eine radiale Reduktion der Inputs ermittelt, dabei jedoch die

Möglichkeiten zur weiteren Reduktion einzelner Inputs, für die ein Slack  $> 0$  vorliegt, außer Acht gelassen. Sollten derartige Potenziale nach radialer Anpassung der entsprechenden Faktoren noch vorliegen, spricht man von partiellen Ineffizienzen. In diesen Fällen kann nicht von einer Effizienz im Sinne von PARETO/KOOPMANS gesprochen werden [HOF2006, S. 41], sondern vielmehr einer „schwachen“ Effizienz im Sinne von DEBREU und FARRELL, wie sie in Unterabschnitt 2.1.2.2 eingeführt wurde.

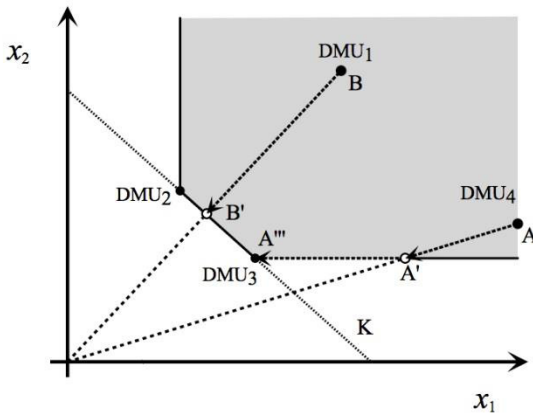


Abbildung 2.21: Slacks im CCR-Modell  
(Quelle: in Anlehnung an [HOF2006, S. 32])

Dieser Problematik tragen CHARNES ET AL. Rechnung, indem sie eine DMU erst dann als CCR-effizient bezeichnen, wenn sie sowohl einen Effizienzwert von 1 als auch für alle Slacks einen Wert von 0 aufweist [LOV1993, S. 14] [RAD2008, S. 37] [HAA2004, S. 59].

Deshalb werden nach der Bestimmung des Effizienzwertes in einem zweiten Schritt die Schlüpfen  $\mathbf{s}_k^- = (s_{k1}^-, \dots, s_{km}^-, \dots, s_{kM}^-)^T \in \mathbb{R}_+^M$  und  $\mathbf{s}_k^+ = (s_{k1}^+, \dots, s_{ks}^+, \dots, s_{kS}^+)^T \in \mathbb{R}_+^S$  der einzelnen In- und Outputs berechnet:

$$\begin{aligned}
& \max \sum_{m=1}^M s_{km}^- + \sum_{s=1}^S s_{ks}^+ \quad (\text{Slack-CCR-I}) \\
& \text{u. d. N.} \quad h_k^* x_{km} = \sum_{j=1}^J \mu_{kj} x_{jm} + s_{km}^-, \quad m = 1, \dots, M \\
& \quad y_{ks} = \sum_{j=1}^J \mu_{kj} y_{js} - s_{ks}^+, \quad s = 1, \dots, S \\
& \quad \mu_{kj} \geq 0 \quad \forall j, s_{km}^-, s_{ks}^+ \geq 0 \quad \forall m, s
\end{aligned}$$

Nimmt man nun die optimale Lösung von Problem (Slack-CCR-I) und berechnet damit die PARETO/KOOPMANS-effiziente Aktivität zur DMU<sub>k</sub>, so erhalten wir

$$(x'_k, y'_k) = (h_k^* x_k - s_k^{*-}, y_k + s_k^{*+}) \quad (\text{PKE-A-I})$$

Bei Erreichung dieser angepassten Zielwerte der Inputs gilt eine DMU als effizient im Sinne der DEA. Weitere Möglichkeiten zum Umgang mit der Problematik der Slacks findet der interessierte Leser unter anderem bei LOVELL [LOV1993], TORGENSEN ET AL. [TFK1996] und CANTNER ET AL. [CKH2007].

Da im weiteren Verlauf dieser Arbeit auch outputorientierte Modelle diskutiert werden, soll an dieser Stelle das radiale outputorientierte Modell (beispielhaft in der Formulierung als primales Problem) zumindest formal eingeführt werden:



$$g_k^* = \max g_k \quad (\text{CCR-O-P})$$

$$\text{u. d. N.} \quad g_k y_k \leq \sum_{j=1}^J v_{kj} y_j$$

$$x_k \leq \sum_{j=1}^J v_{kj} x_j$$

$$v_{kj} \geq 0 \quad \forall j$$

Um die Aktivität einer  $DMU_k$  auf PARETO/KOOPMANS-Effizienz outputorientiert zu testen, müssen analog zur inputorientierten Vorgehensweise nach dem Lösen von (CCR-O-P) die Schlüpfе bestimmt werden. Wir verzichten hier auf eine Formulierung der outputorientierten Aufgabe (Slack-CCR-O) und die Darstellung einer entsprechenden (PKE-A-O).

Trotz der großen Akzeptanz und einer Vielzahl von theoretischen wie empirischen Anwendungen mussten sich die CCR-Modelle der Kritik stellen, dass die Annahme konstanter Skalenerträge den meisten Praxisfällen nicht gerecht wird. Insbesondere die in vielen Branchen mit der Produktionsmenge drastisch sinkenden Stückkosten, Preisvorteile beim Einkauf größerer Mengen sowie Vorteile der möglichen Spezialisierung einzelner Arbeitskräfte bei Ausweitung der Produktion forderten ein entsprechendes Modell auf Basis der Annahme variabler Skalenerträge [HAA2004, S. 52].

### 2.2.3 BCC-Modelle mit Input-/Outputorientierung

Die BCC-Modelle gehen zurück auf den vielbeachteten Artikel von BANKER, CHARNES UND COOPER aus dem Jahr 1984 [BCC1984] und sind als Weiterentwicklung der im vorigen Abschnitt behandelten CCR-Modelle zu betrachten. Der wesentliche Unterschied liegt in der Tatsache, dass die BCC-Modelle die An-

nahme nicht konstanter, also variabler, Skalenertragslagen zulassen. In Unterabschnitt 2.1.2.2 wurde der Begriff der **Skalenertragslage** bereits in seinen Grundlagen eingeführt. Vor der konkreten Darstellung des BCC-Ansatzes sollen die Frage der Skalenertragslage vertieft und die resultierenden Implikationen für die Technologiemenge herausgearbeitet werden.

Konkret macht sich die Skalenertragslage fest an der Skalanelastizität, die sich als Relation von Steigung und Produktivität definiert.

Abbildung 2.22 zeigt eine Produktionsfunktion mit konstanten Skalenerträgen (CRS) und eine ertragsgesetzliche Produktionsfunktion, die naturgemäß auf einer variablen Skalenertragslage (VRS) basiert:

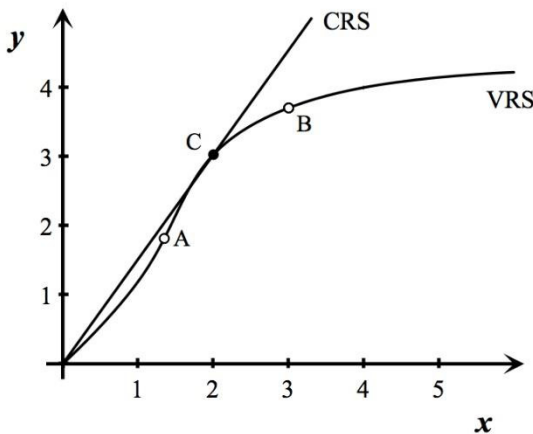


Abbildung 2.22: Produktionsfunktionen unter konstanten und variablen Skalenertragslagen  
(Quelle: eigene Darstellung)

Der Punkt C entspricht der DMU 2 in den Abbildungen 2.13 und 2.14. Wie bereits im Nachgang zu Abbildung 2.13 erwähnt wurde, hat er höchste Produktivität. Allgemein versteht man unter dem Skalenertrag die Steigung in Relation zur Produktivität. Sie hat also im Punkt C den Wert 1; Punkt A hat einen Skalenertrag  $>$

1, wohingegen Punkt B einen solchen  $< 1$  aufweist. Hierzu haben sich die Begriffe steigender bzw. fallender Skalenerträge etabliert.

Die Notwendigkeit der Berücksichtigung variabler Skalenerträge bedarf keiner komplizierten Begründung, da in der Realität in kaum einem Szenario von durchweg konstanter Skalenertragslage ausgegangen werden kann. ABBOT UND DOUCOLIAGOS bringen die praktische Relevanz variabler Skalenerträge am Beispiel von Universitäten auf den Punkt: „A university may be technically efficient but it may still be producing too little or too much output. This information is derived from measures of scale efficiency. Scale efficiency is the extent to which an institution can take advantage of returns to scale by altering its size toward the optimal size (defined as the region in which there are constant returns to scale in the relationship between outputs and inputs).“ [ABD2003, S. 91]. Dennoch sollte daraus nicht der Schluss gezogen werden, dass grundsätzlich von variablen Skalenertragslagen auszugehen und folglich immer ein BCC-Modell anzuwenden sei. Vielmehr muss im Rahmen der DEA-Modellierung die Frage gestellt werden, ob Größenunterschiede zwischen den betrachteten DMUs Teil der Betrachtung sein oder eben gezielt ausgeklammert werden sollen. Die verschiedenen Aspekte dieser Fragestellung werden in Unterkapitel 4.4 ausführlicher behandelt. Zum jetzigen Zeitpunkt soll lediglich festgehalten werden, dass BCC-Modelle durch die Unterstellung variabler Skalenerträge die Größenunterschiede zwischen den einzelnen DMUs in der Untersuchung berücksichtigen und gewissermaßen ausgleichen: In BCC-Modellen können auch solche Unternehmen als effizient bewertet werden, die sich nicht im Bereich der optimalen Unternehmensgröße befinden, und dadurch Produktivitätspotenziale nicht vollständig heben.

Die veränderte Skalenertragsannahme hat auch Konsequenzen bezüglich der **Technologie**. Die in Unterabschnitt 2.1.2.1 genannten Axiome 1 bis 4 haben weiterhin Geltung. Die Abgrenzung vom vorab erläuterten Modell mit konstanten Skalenerträgen äußert sich insbesondere im Entfall des dort geltenden Axioms 5 (a) zugunsten des nun geltenden Axioms 5 (d). Die daraus resultierende BCC-Technologie ist in Abbildung 2.23 hellgrau dargestellt. Die dunkelgrauen Flächen kennzeichnen die Bereiche, die zwar zu einer entsprechenden CCR-Technologie gehören, nicht jedoch zur BCC-Technologie.

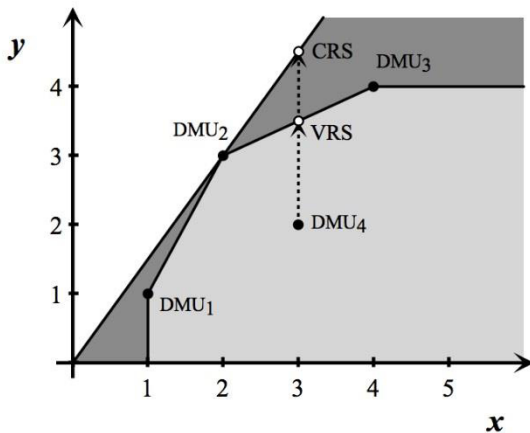


Abbildung 2.23: Technologie in BCC- und CCR-Modellen  
(Quelle: eigene Darstellung)

Es ist klar ersichtlich, dass die BCC-Technologie eine Teilmenge der CCR-Technologie ist. Daraus folgt, dass eine unter konstanten Skalenerträgen effiziente Aktivität bei identischem Daten-Set auch unter variablen Skalenerträgen effizient ist; da der Technologieraum des CCR-Modells im BCC-Modell lediglich verkleinert wird, liegt eine unter CRS effiziente Aktivität auch unter VRS auf dem effizienten Rand. In BCC-Modellen effiziente Aktivitäten müssen in entsprechenden CCR-Modellen jedoch keineswegs effizient sein. Durch den verkleinerten Technologieraum werden Aktivitäten im BCC-Modell, verglichen mit der CCR-Technologie, tendenziell eher als effizient eingestuft. Dadurch, dass der effiziente Rand enger an die einzelnen DMUs rückt oder die Distanz zumindest gleich bleibt, kann davon ausgegangen werden, dass die Effizienzwerte der einzelnen DMUs unter variablen Skalenerträgen immer größer oder gleich sind, verglichen mit den entsprechenden CCR-Werten [CRO2005]. Abbildung 2.23 zeigt am (outputorientierten) Beispiel der unter beiden Technologien ineffizienten  $DMU_4$ , dass sich der Abstand zum effizienten Rand (gestrichelt dargestellt) bei

dem Wechsel von der CCR-Technologie zur BCC-Technologie verkleinert und der entsprechende Effizienzwert folglich höher ausfallen muss.

Das Effizienzmaß im CCR-Modell enthält sowohl Effekte der technischen als auch der Skaleneffizienz. So ist beispielsweise die DMU<sub>2</sub> in Abbildung 2.23 sowohl technisch als auch skaleneffizient. Bei Anwendung eines BCC-Modells erhält man hingegen ein Maß zur reinen Beurteilung der technischen Effizienz, also bereinigt um den Einfluss der Skaleneffizienz [CKH2007, S. 159]. Auf dieser Basis lässt sich der reine Einfluss der Betriebsgröße auf die Effizienz, die Skaleneffizienz (SE), aus der Relation der Effizienzergebnisse der CCR- und BCC-Modelle gewinnen:

$$SE = \frac{eff_{CCR}}{eff_{BCC}}$$

Analog zur eingeführten Notation und der Darstellung der CCR-Modelle lautet die inputorientierte Variante des BCC-Modells in der **Envelopmentform**:

$$\bar{h}_k^* = \min h_k \quad (\text{BCC-I-E})$$

$$\text{u. d. N.} \quad h_k \mathbf{x}_k \geq \alpha_k \sum_{j=1}^J \lambda_{kj} \mathbf{x}_j$$

$$\mathbf{y}_k \leq \alpha_k \sum_{j=1}^J \lambda_{kj} \mathbf{y}_j$$

$$\sum_{j=1}^J \lambda_{kj} = 1$$

$$\alpha_k = 1, \lambda_{kj} \geq 0 \quad \forall j$$

Dementsprechend erhält man (analog zum CCR-Modell) für eine effiziente  $DMU_k$  ein  $\bar{h}_k^* = 1$  und für eine ineffiziente  $DMU_k$  ein  $\bar{h}_k^* < 1$ .

Mit der jetzt eingeführten Symbolik berechnet sich die inputorientierte Skaleneffizienz zu:

$$SE = \frac{h_k^*}{\bar{h}_k^*} \quad (SE-I)$$

Während unter konstanten Skalenerträgen zwischen den Effizienzwerten der input- und outputorientierten Betrachtung ein reziprokes Verhältnis besteht, können die Effizienzwerte in Modellen variabler Skalenerträge je nach Orientierung voneinander abweichen. Dies gilt jedoch nur für die Höhe der Effizienzwerte, nicht für die Frage ob eine Einheit effizient ist oder nicht [CKH2007, S. 207 f.].

Die Variablentransformation  $\mu_{kj} = \alpha_k \lambda_{kj}$  überführt die nichtlineare Aufgabe (BCC-I-E) in eine lineare, das primale Problem (BCC-I-P):

$$\bar{h}_k^* = \min h_k \quad (BCC-I-P)$$

$$\text{u. d. N.} \quad h_k x_k \geq \sum_{j=1}^J \mu_{kj} x_j$$

$$y_k \leq \sum_{j=1}^J \mu_{kj} y_j$$

$$\sum_{j=1}^J \mu_{kj} = 1$$

$$\mu_{kj} \geq 0 \quad \forall j$$

Durch Dualisierung ergibt sich die **Multiplifierform** (BCC-I-M):

$$\bar{h}_k^* = \max \mathbf{U}_k^T \mathbf{y}_k + u_k \quad (\text{BCC-I-M})$$

$$\text{u. d. N.} \quad \mathbf{V}_k^T \mathbf{x}_k = 1$$

$$\mathbf{U}_k^T \mathbf{y}_j + u_k - \mathbf{V}_k^T \mathbf{x}_j \leq 0, \quad j = 1, \dots, J$$

$$\mathbf{U}_k, \mathbf{V}_k \geq \mathbf{0}, u_k \text{ frei}$$

Analog zum CCR-Modell in Abschnitt 2.2.2 lässt sich die Aufgabe (BCC-I-M) ebenfalls durch eine Variablentransformation in ein **Quotientenproblem** überführen; auf eine konkrete Darstellung wird an dieser Stelle verzichtet.

Das Vorzeichen des optimalen  $u_k$  im Optimierungsproblem (BCC-I-M) gibt Auskunft über die Skalenertragslage, wobei im Fall eines freien  $u_k$  von variablen Skalenerträgen gesprochen wird, wie wir es für das BCC-Modell definiert haben. Liegt nun die optimale Lösung  $\mathbf{U}_k^*, \mathbf{V}_k^*, u_k^*, \bar{h}_k^*$  vor, so lässt sich anhand der konkreten Ausprägung des  $u_k^*$  die Skalenertragslage der jeweiligen DMU interpretieren: Bei einem  $u_k^* = 0$  liegen konstante Skalenerträge vor, was allein durch einen Vergleich mit der Gleichung (CCR-I-M) augenscheinlich wird. Im Fall eines  $u_k^* > 0$  spricht man von steigenden und analog für  $u_k^* < 0$  von fallenden Skalenerträgen.

Hierbei ist zu beachten, dass ein errechnetes  $u_k^*$  und die dazugehörigen Optimalgewichte keineswegs eindeutig sein müssen: Das ermittelte  $u_k^*$  steht in enger Beziehung zu dem  $\mathbf{y}$ -Achsenabschnitt einer Effizienzisoquante (zum Begriff der Effizienzisoquante vgl. [DEL2015]) der DMU<sub>k</sub> [STE2002, S. 28 f.]. Abbildung 2.24 zeigt am Beispiel der DMU<sub>1</sub> eine Schar von Effizienzisoquanten, die jedoch mit jeweils unterschiedlichen  $u_k^*$  einhergehen:

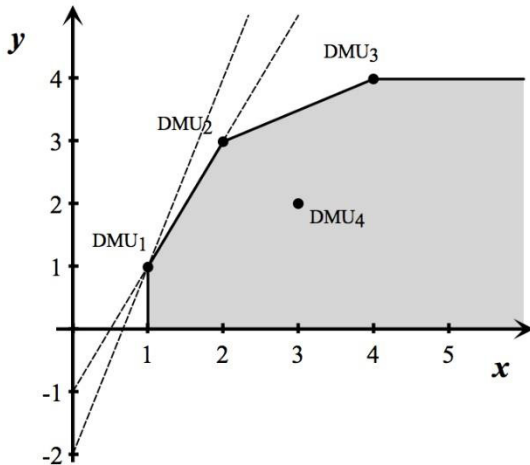


Abbildung 2.24: Schar von Effizienziisoquanten  
(Quelle: [ROD2011, S. 18])

Die unterschiedlichen Skalenertragslagen einzelner DMUs werden am bekannten Beispiel-Set von DMUs in Abbildung 2.25 verdeutlicht. Für die Aktivität von  $DMU_2$  verläuft die Effizienziisoquante (durchgehende Linie) durch den Ursprung, folglich liegen hier konstante Skalenerträge vor. Die Isoquante zu  $DMU_1$  (weit gestrichelte Linie) schneidet die  $y$ -Achse in -1, arbeitet unter steigenden Skalenerträgen, und  $DMU_3$  analog (eng gestrichelte Linie) mit Schnittpunkt der  $y$ -Achse bei 2 unter fallenden Skalenerträgen.



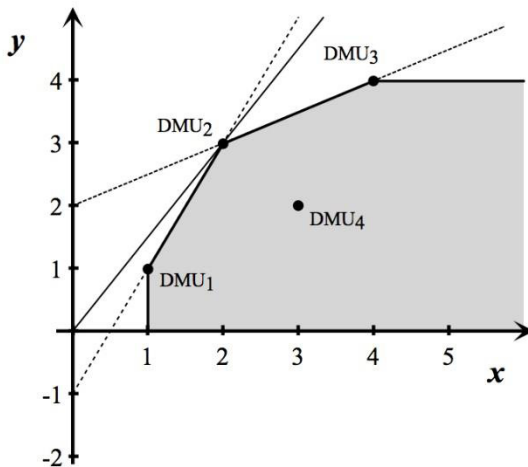


Abbildung 2.25: Effizienziisoquanten und  $y$ -Achsenabschnitte  
(Quelle: in Anlehnung an [STE2002, S. 28])

Die Thematik der **Slacks** wurde im Rahmen der CCR-Modelle bereits behandelt und hat bei den BCC-Modellen gleichermaßen Geltung. Dementsprechend kann auch erst von BCC-Effizienz einer Einheit gesprochen werden, wenn diese sowohl einen Effizienzwert von 1, als auch für alle Slacks einen Wert von 0 aufweist:

$$\begin{aligned}
& \max \sum_{m=1}^M s_{km}^- + \sum_{s=1}^S s_{ks}^+ && \text{(Slack-BCC-I)} \\
\text{u. d. N.} \quad & \bar{h}_k^* x_{km} = \sum_{j=1}^J \mu_{kj} x_{jm} + s_{km}^-, && m = 1, \dots, M \\
& y_{ks} = \sum_{j=1}^J \mu_{kj} y_{js} - s_{ks}^+, && s = 1, \dots, S \\
& \sum_{j=1}^J \mu_{kj} = 1 \\
& \mu_{kj} \geq 0 \quad \forall j, s_{km}^-, s_{ks}^+ \geq 0 \quad \forall m, s
\end{aligned}$$

CANTNER ET AL. weisen darauf hin, dass unter Verwendung von Modellen mit variablen Skalenerträgen bei inputorientierten Untersuchungen auch Output-Slacks auftreten können und umgekehrt [CKH2007, S. 225]. Es bedarf im Rahmen von BCC-Modellen grundsätzlich einer Überprüfung der In- und Output-Slacks. Zu der hier dargestellten zweistufigen Vorgehensweise zum Auffinden PARETO/KOOPMANS-effizienter Lösungen gibt es bekanntlich eine einstufige Form, deren Dualisierung sich in einer strikten Positivitätsforderung der Gewichte niederschlägt [BCC1984]. Im weiteren Verlauf dieser Arbeit wird von solchen Feinheiten abgesehen.

Wie bereits im Rahmen der CCR-Modelle sei auch an dieser Stelle die outputorientierte Modellvariante in Form des primalen Problems beigelegt:

$$\bar{g}_k^* = \max g_k \quad (\text{BCC-O-P})$$

$$\text{u. d. N.} \quad g_k y_k \leq \sum_{j=1}^J v_{kj} y_j$$

$$x_k \geq \sum_{j=1}^J v_{kj} x_j$$

$$\sum_{j=1}^J v_{kj} = 1$$

$$v_{kj} \geq 0 \quad \forall j$$

## 2.2.4 Zeitlich-dynamische DEA-Modelle

Bei den bisher vorgestellten Effizienzbegriffen und -maßen handelt es sich um rein statische Betrachtungen, die auf Basis einer Momentaufnahme arbeiten und auf einen zurückliegenden Zeitraum bezogene Effizienzaussagen tätigen. Eine zeitlich-dynamische Betrachtung von Effizienzen kann in mehreren Hinsichten von Interesse sein und wertvolle Erkenntnisse liefern.

Dem stehen allerdings ein vergleichsweise hoher Anspruch an die Datenbasis sowie ein Mehraufwand bei Modellierung und Berechnung entgegen. So lassen sich beispielsweise relative Effizienzwerte einer DMU aus verschiedenen Perioden nicht ohne weiteres in Relation zueinander setzen, eben da es sich um relative Effizienzen handelt. Die maßgebliche Bezugsgröße, also der effiziente Rand der (approximierten) Produktionsfunktion, hängt von den in der betreffenden Periode vorherrschenden Produktionsbedingungen ab und kann keineswegs als konstant angenommen werden [CKH2007, S. 247]. Insofern bedarf es weiterer Maßnahmen, um die Effizienzergebnisse verschiedener Perioden sinnvoll in Relation setzen zu können – schon deshalb muss abgewogen werden, ob Anforderungen und Nutzen in einem positiven Verhältnis stehen. Die konkrete Frage nach Einsatz und Modellierung zeitlich-dynamischer DEA-Modelle wird jedoch nicht Thema

dieser Arbeit sein, die beiden etabliertesten dynamischen Methoden werden auf den folgenden Seiten lediglich in ihren Grundzügen vorgestellt.

#### 2.2.4.1 *MALMQUIST-Produktivitätsindex*

Eine der populärsten Methoden der dynamischen Effizienzanalyse ist der MALMQUIST-Index, der auf das Jahr 1953 zurückgeht [MAL1953]. Als MALMQUIST-Index wird in der Literatur eine Indexzahl dann bezeichnet, wenn sie Distanzfunktionen über einen Quotienten in eine Beziehung setzt [CKH2007, S. 250]; erst knapp 30 Jahre später verwendeten CAVES ET AL. diesen Index zur Beschreibung von Produktivitäten [CCD1982]. Dabei wird eine sinnvolle Darstellung der Effizienzentwicklung einer DMU in zwei verschiedenen Perioden dadurch ermöglicht, dass mithilfe des MALMQUIST-Index neben den Effizienzänderungen auch die Veränderungen des effizienten Randes zwischen den Perioden berücksichtigt werden können.

Dementsprechend leitet sich ein Effizienzwert nach MALMQUIST aus zwei Komponenten ab: Der „Catch-up“ steht für die Veränderung des relativen Effizienzwertes einer DMU zwischen zwei Perioden, während der „Frontier-shift“ die Veränderung der Produktionsmöglichkeiten (also des Technologieraums) erfasst [ALF2008, S. 53 f.]. Anhand von Abbildung 2.26 lassen sich beide Komponenten anschaulich am Beispiel einer inputorientierten Betrachtung nachvollziehen. Dabei habe eine DMU in Periode  $t_0$  die Aktivität D und in Periode  $t_1$  die Aktivität G:

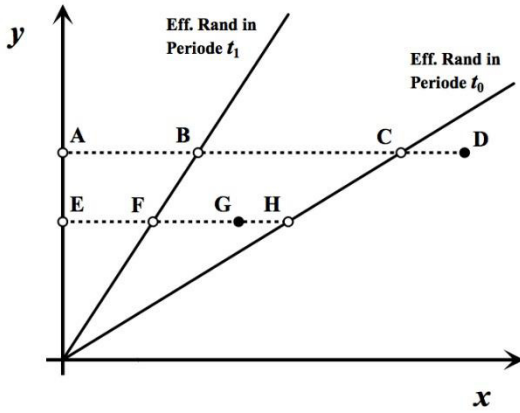


Abbildung 2.26: Zeitliche Dynamik einer Produktionsfunktion  
(Quelle: in Anlehnung an [TON2004, S. 205])

Der Catch-up  $\Omega$  setzt die relative Effizienz der betrachteten DMU in Periode  $t_0$  in Relation zum analogen Wert in Periode  $t_1$ , also konkret:

$$\Omega = \frac{|EF|/|EG|}{|AC|/|AD|}$$

Folglich bedeutet beispielsweise ein  $\Omega < 1$  ein Absinken der technischen Effizienz in Periode  $t_1$  im Vergleich zu Periode  $t_0$ .

Der Frontier-shift ist für die betrachteten Perioden einzeln zu ermitteln. Zur Darstellung der Verschiebung der Produktionsmöglichkeiten bezogen auf die Aktivität der DMU in Periode  $t_0$ :

$$\Phi_1 = \frac{|AC|/|AD|}{|AB|/|AD|}$$

Analog für die Aktivität in Periode  $t_1$ :

$$\Phi_2 = \frac{|EH|/|EG|}{|EF|/|EG|}$$

CAVES ET AL. [CCD1982] kamen über diese beiden periodenbezogenen Betrachtungen auch zu zwei verschiedenen MALMQUIST-Indizes, die auf je eine Periode als Bezug für die maßgebliche Technologie abstellten. Später entwickelten FÄRE ET AL. [FGL1992] [FGL1994] [FGN1994] das Modell dahingehend weiter, dass sie über eine Aggregation der beiden genannten Frontier-shift-Komponenten den Weg frei machten für eine Betrachtung unter Berücksichtigung der Produktionsmöglichkeiten beider Perioden. Diese Aggregation der Frontier-shifts beider Perioden lautet in Fortführung des graphischen Beispiels:

$$\Phi = \sqrt{\Phi_1 \cdot \Phi_2}$$

Werte von  $\Phi > 1$  sind als technologischer Fortschritt zu interpretieren und bedeuten eine Ausweitung der Technologiemenge. Dieser Effekt liegt in Abbildung 2.26 vor und ist deutlich an der Verschiebung der Produktionsmöglichkeiten von  $t_0$  zu  $t_1$  zu erkennen. Dementsprechend müssen DMUs in diesem Fall ihre Produktivität steigern, um ihre relative Effizienz aus Periode  $t_0$  in der Periode  $t_1$  halten zu können.

Der MALMQUIST-Index errechnet sich nun durch Multiplikation der beiden dargestellten Komponenten, des Catch-up und des Frontier-shift. Plastisch gesprochen wird die Effizienzveränderung einer DMU um den enthaltenen Einfluss der im zeitlichen Verlauf veränderten Produktionsbedingungen bereinigt:

$$MI = \Omega \cdot \Phi$$

Ein  $MI > 1$  gibt eine Erhöhung der totalen Faktorproduktivität der betrachteten DMU in  $t_1$  an, bezogen auf die Vergleichsperiode  $t_0$ .

Grundsätzlich ist der MALMQUIST-Index nur in Zusammenhang mit Modellen konstanter Skalenerträge sinnvoll anwendbar, weil es im Rahmen von variablen Skalenertragsannahmen zu Fehlinterpretationen kommen kann [OUV2004, S. 759]. Zwar gab es verschiedentliche Ansätze für einen validen Einsatz des MALMQUIST-Index unter variablen Skalenerträgen, doch konnte bislang keiner davon eine breite Akzeptanz in der Fachwelt erreichen; bis heute wird der MALMQUIST-Index in erster Linie unter konstanten Skalenerträgen eingesetzt. Der interessierte Leser sei zur Vertiefung auf CANTNER ET AL. [CKH2007, S. 262

f.], COELLI ET AL. [CRO2005, S. 293] und TONE [TON2004, S. 215 ff.] verwiesen.

#### 2.2.4.2 *Window-Analysis*

Der Window-Analysis, die 1985 von CHARNES ET AL. [CCC1985] im Zusammenhang mit der DEA eingeführt wurde, liegt die Idee zu Grunde, die relative Effizienz einer DMU nicht nur über die Produktivitäten der Vergleichs-DMUs zu bestimmen, sondern die Effizienzen aller betrachteten DMUs über alle relevanten Perioden in die Betrachtung einzubeziehen [ALF2008, S. 49]. Somit werden die Aktivitäten verschiedener Perioden einer DMU in der Effizienzberechnung als die Aktivitäten verschiedener DMUs betrachtet [CCG1994, S. 153] oder anders formuliert: Eine DMU wird nicht als „Entscheidungseinheit“ definiert, sondern als „Entscheidungseinheit einer bestimmten Periode“ [RES2010, S. 10].

Abbildung 2.27 zeigt einen Auszug aus der Datenbasis des genannten Artikels von CHARNES ET AL., der 14 Serviceeinheiten der U.S. Air Force über einen Zeitraum von sieben Perioden, P1, ..., P7 untersucht. Zugunsten einer besseren Übersichtlichkeit wird hier eine auf die DMUs A, ..., H verkürzte tabellarische Darstellung gewählt, um das Prinzip der Window-Analysis zu verdeutlichen. Der Begriff „Window“ leitet sich aus der Tatsache ab, dass im Rahmen der Window-Analysis nicht alle Perioden gleichzeitig in die Effizienzberechnung eingehen. Es wird vielmehr im Sinne eines Fensters eine feste Anzahl von Perioden definiert, die in diesem Verfahren gleichzeitig in eine Effizienzberechnung einfließen. CHARNES ET AL. definierten dieses Fenster auf genau drei Perioden, in Abbildung 2.27 beispielhaft über die Perioden P1, P2 und P3 anhand einer gestrichelten Umrahmung dargestellt. Um dennoch alle sieben Perioden in der Untersuchung zu berücksichtigen, wird dieses Fenster von drei Perioden so lange um je eine Periode verschoben (In Abbildung 2.27 durch den überlagernden Pfeil angedeutet), bis es die letzte Periode einschließt. Im dargestellten Beispiel ergeben sich dementsprechend fünf Berechnungen über je drei Perioden. Innerhalb dieser fünf Berechnungen werden alle Aktivitäten im jeweiligen Fenster betrachtet, als seien sie

Aktivitäten einer identischen Periode und von verschiedenen DMUs; folglich erhöht sich die Anzahl der DMUs in einer Berechnung bei einem Fenster von drei Perioden auch um den Faktor drei. Diesen Effekt machen sich CHARNES ET AL. gezielt zunutze: Ein mathematisch ungünstiges Verhältnis zwischen der Anzahl der DMUs und der In- und Outputs wird durch den Einsatz der Window-Analyse automatisch korrigiert [SCH1994, S. 184] [PVY2004, S. 372]. Der Autor wird diese Problematik im Rahmen der später folgenden DEA-Modellierung gesondert thematisieren.

Die Effizienzergebnisse der einzelnen Fenster-Untersuchungen finden sich in der Abbildung 2.27 jeweils in einer separaten Zeile der einzelnen DMUs. Dementsprechend kann aus den Zeilen der DMUs die Entwicklung bzw. ein Trend der zeitlichen Veränderung der Effizienzwerte der einzelnen DMUs abgelesen werden. Die Betrachtung der Spalten gibt einen Überblick über den Vergleich der Effizienzwerte der verschiedenen DMUs. In den nicht am Rand liegenden Spalten, in unserem Beispiel die Perioden P2, ..., P6, liegt für jede DMU mehr als ein Effizienzwert vor, so dass sich hier auch Aussagen zur Stabilität bzw. Validität der mit der Window-Analysis gewonnen Ergebnisse ableiten lassen [CST2002, S. 273].

Ein Kritikpunkt an der Window-Analysis liegt in der Frage der Größe des rollierenden Untersuchungsfensters. Sowohl ein zu großes als auch ein zu kleines Fenster kann unter bestimmten Umständen negative Konsequenzen haben, dennoch gibt es bislang keine etablierte Methode zur Bestimmung der optimalen Fenstergröße [PVY2004, S. 373] [CSZ2004, S. 27]; BÜRKLE UND RABER empfehlen, die Entscheidung über die Größe der Fenster auf inhaltlicher Basis zu treffen, sich also an der praktischen Ausgangslage zu orientieren [BUR1995, S. 5]. REUCHER UND SARTORIUS gehen in diesem Zusammenhang auf die Auswirkungen einer Veränderung der Technologie ein: Sollte sich innerhalb des für die Größe des Betrachtungsfensters definierten Zeitraums eine signifikante Veränderung der Technologie (zum Beispiel durch technologischen Fortschritt) ergeben, würde das zu ungewollten Verzerrungen der Effizienzergebnisse führen. Deshalb sollte der Zeitraum der Fenstergröße so kurz gewählt werden, dass darin keine nennenswerten Veränderungen der Technologie zu verzeichnen sind [RES2010, S. 10 f.].



DMU	P1	P2	P3	P4	P5	P6	P7
<b>A</b>	97,89	97,31 97,36	98,14 97,53 96,21	97,04 95,92 95,79	94,54 94,63 94,33	97,64 97,24	97,74
<b>B</b>	93,90	95,67 96,72	96,14 96,42 95,75	94,63 94,14 94,54	93,26 93,46 93,02	96,02 96,02	94,49
<b>C</b>	93,77	91,53 91,77	95,26 95,55 93,21	94,29 95,04 93,20	94,83 93,09 93,59	92,21 92,32	92,83
<b>D</b>	99,72	96,15 97,91	95,06 95,70 94,79	100,00 100,00 99,71	94,51 94,39 94,95	94,76 94,67	89,37
<b>E</b>	100,00	100,00 100,00	100,00 100,00 98,97	100,00 99,05 99,37	100,00 100,00 100,00	100,00 100,00	100,00
<b>F</b>	97,42	93,48 93,60	96,07 96,24 94,46	93,56 91,75 91,73	92,49 92,32 92,68	92,35 91,98	99,64
<b>G</b>	90,98	92,80 93,67	95,96 96,80 93,34	99,52 94,48 91,94	91,73 89,79 89,35	95,58 95,14	96,38
<b>H</b>	100,00	100,00 100,00	100,00 100,00 100,00	100,00 100,00 100,00	100,00 100,00 100,00	100,00 100,00	100,00

Abbildung 2.27: Beispiel einer Window-Analyse  
(Quelle: in Anlehnung an [CCC1985, S. 104])

Auch die Problematik einer Behandlung mehrerer Aktivitäten einer DMU als Aktivitäten verschiedener DMUs muss kritisch hinterfragt werden.

## 2.2.5 Weitere Modellvarianten und Modellerweiterungen

Wie sich in Kapitel 3 zeigen wird, ist die DEA nunmehr seit Jahrzehnten Gegenstand einer Vielzahl empirischer, vor allem aber auch theoretischer Veröffentlichungen. Dabei wurde das Konzept der DEA stets weiterentwickelt und um unterschiedlichste Komponenten und flankierende Modelle erweitert. Im vorliegenden Abschnitt werden aus der schier unendlichen Fülle der Modellvarianten und -erweiterungen nur die Wichtigsten vorgestellt, um dem Leser einen Überblick über die etablierte DEA zu ermöglichen. Da diese nur teilweise im weiteren Verlauf dieser Arbeit wieder aufgegriffen werden, fällt auch der Umfang ihrer Darstellung unterschiedlich aus.

### 2.2.5.1 *Nicht-orientierte Modelle*

Während die bisher eingeführten CCR- und BCC-Modelle entweder als input- oder outputorientiert definiert wurden, zeichnen sich die nicht-orientierten Modelle durch eine gleichzeitige Betrachtung der In- und Outputseite aus. Das **Additive Modell** leitet sich aus einer Arbeit von CHARNES & COOPER aus dem Jahr 1961 [CHC1961] ab und wurde 1985 von CHARNES ET AL. [CCG1985] vorgestellt; über eine Maximierung der In- und Outputschlüpfе errechnet das Additive Modell die Ineffizienz der betreffenden DMU, was neben der angesprochenen nicht-orientierten Betrachtung auch isoliert bezüglich der In- oder Outputschlüpfе erfolgen kann.

Abbildung 2.28 stellt die Möglichkeiten einer ineffizienten  $DMU_4$  zur Verbesserung ihrer Effizienz dar:

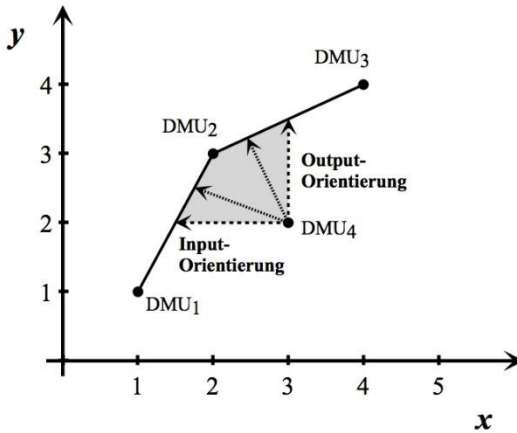


Abbildung 2.28: Graphische Grundlagen nicht-orientierter DEA-Modelle am Beispiel einer VRS-Technologie  
(Quelle: eigene Darstellung)

Neben den bereits bekannten Varianten der Inputreduzierung und Outputerweiterung (weit gestrichelt) sind hier zusätzlich beispielhafte Pfeile einer gleichzeitigen Anpassung von In- und Outputs dargestellt (eng gestrichelt). Die Pfeile entsprechen einer vektorbasierten Darstellung der genannten Schlüpfе im Fall einer In- oder Outputbetrachtung (weit gestrichelt) oder einer nicht-orientierten Betrachtung (eng gestrichelt); erst wenn das maximierte Effizienzmaß null entspricht, existieren keine Schlüpfе mehr und die DMU liegt auf dem effizienten Rand. Neben dem in Abbildung 2.28 dargestellten Fall einer VRS-Technologie kann das Additive Modell auch unter konstanten Skalenerträgen Anwendung finden [SCH1994, S. 176]. Eine Darstellung von Erweiterungen und Abwandlungen des Additiven Modells liefern z. B. SCHEEL [SCH2000, S. 106 ff.] und COOPER ET AL. [CPP1999].

**Multiplikative Modelle** werden hier als Bestandteil der nicht-orientierten Modelle eingeordnet; sie maximieren ebenfalls die Schlüpfе, legen jedoch logarithmierte Inputs und Outputs zu Grunde. BANKER UND MAINDIRATTA [BAM1988] haben zwar nachgewiesen, dass sich multiplikative Modelle auch in orientierte DEA-Modelle integrieren lassen, doch haben sich derartige Varianten nicht durchsetzen können [SCH1994, S. 178]. Auch insgesamt blieben Multiplikative Modelle in der Praxis eine Randerscheinung [BCS2004, S. 62]; zu den ausführlichen Grundlagen sei neben den bereits angeführten Quellen auf CHARNES ET AL. 1982 [CCS1982] und 1983 [CCS1983] verwiesen. Sowohl Additive als auch Multiplikative Modelle spielen in unseren weiteren Überlegungen keine Rolle.

#### *2.2.5.2 Unkontrollierbare Variable und Variablen-Begrenzungen*

Bei der Untersuchung von Effizienzen liegt die Zielsetzung im Normalfall in der Steigerung derselben. Dies geschieht über die Reduktion der Inputs und/oder die Ausweitung der Outputs. Um dem Ziel der Untersuchung gerecht zu werden, gilt es, die relevanten Einflussfaktoren zu identifizieren und in die Berechnungen einfließen zu lassen. Ein Sonderfall liegt dabei in sogenannten unkontrollierbaren Variablen, die das Effizienzergebnis zwar durchaus beeinflussen, jedoch nicht oder nur in begrenztem Maße durch die DMUs gesteuert werden können, wie z. B. das Wettbewerbsumfeld, das Konsumentenverhalten oder die Mechanismen der relevanten Märkte [HAN1999, S. 133].

Den ersten Ansatz zur Modellierung solcher Fälle lieferten BANKER UND MOREY 1986 [BMO1986] am Beispiel von 60 Fast-Food-Restaurants. Dabei nehmen sie die als unkontrollierbar identifizierten In- bzw. Outputs zwar als konstant an, nehmen sie jedoch nicht vollständig aus der Berechnung heraus. Die unter Berücksichtigung unkontrollierbarer Variabler ermittelten Effizienzwerte der einzelnen DMUs sind tendenziell höher als bei Unterstellung ausschließlich voll steuerbarer In- und Outputs. Die Einstufung einer DMU als effizient oder ineffizient bleibt davon jedoch unberührt [SCH2000, S. 151]. HAAS legt ergänzend dar, dass die Bestimmung der Slacks auch für unkontrollierbare Variable einen Mehrwert lie-

fert, da sich hier die (unter gegebenen Bedingungen) nicht zu hebenden Potenziale einer Inputreduktion bzw. Outputerhöhung zeigen [HAA2002, S. 65]. Trotz dieses pragmatischen Lösungsansatzes zum Umgang mit unkontrollierbaren Variablen weist HOFFMANN zu Recht darauf hin, dass ihr Anteil begrenzt bleiben muss, damit noch von einer ausreichenden Diskriminanzstärke des Modells ausgegangen werden kann [HOF2006, S. 70].

Auch der Fall begrenzter Variabler ergibt sich zwangsläufig aus der Praxis, zum Beispiel da gewisse Faktoren naturgemäß eine gewisse Mindest- oder Höchstmenge im Rahmen der Transformation nicht über- bzw. unterschreiten können oder dürfen. Solche Einschränkungen können durch die Einführung von oberen oder unteren Schranken für einzelne oder aber die Summe aller Intensitäten in der Envelopment-Form Berücksichtigung finden [KLE2002, S. 145ff.]. In der Multiplier-Form können Einschränkungen der Gewichte, in der einschlägigen Literatur bekannt als „Weight-restrictions“, relativ oder absolut vorgenommen werden. Tiefere Einblicke liefern THOMPSON ET AL. [TST1986], CHARNES ET AL. [CCH1990] und COOPER ET AL. [CST2006, S. S. 165 ff.].

### 2.2.5.3 *Super-Effizienz*

Die Super-Effizienzmaße richten ihren Fokus ausschließlich auf die in den bisher vorgestellten Verfahren als effizient eingestuften DMUs. ANDERSEN UND PETERSEN lieferten 1993 einen Ansatz, der eine Differenzierung der effizienten Einheiten ermöglicht: Dazu werden diese effizienten Einheiten jeweils einzeln aus der Effizienzberechnung herausgenommen und danach der Abstand der isolierten Einheit zu dem neuen effizienten Rand der Technologie betrachtet [ANP1993].

Abbildung 2.29 zeigt am Beispiel eines Modells mit variablen Skalenerträgen, dass  $DMU_2$  in der ursprünglichen Effizienzbetrachtung auf dem effizienten Rand liegt. Wenn nun diese DMU aus der Betrachtung herausgenommen wird, verkleinert sich die Technologie um die hellgraue Fläche und der effiziente Rand verschiebt sich exakt auf die Grenze zwischen dem hellgrauen und dem

dunkelgrauen Bereich. Der gestrichelte Pfeil zeigt den Abstand der  $DMU_2$  zu diesem neuen effizienten Rand:

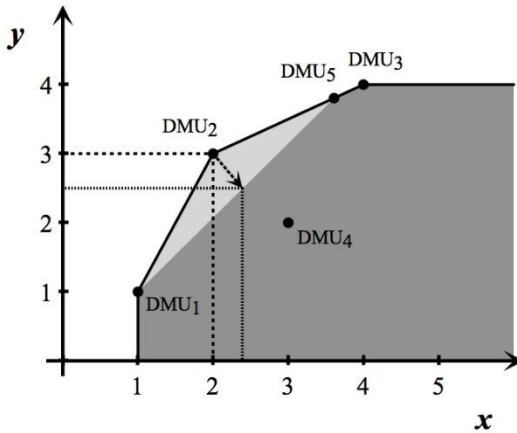


Abbildung 2.29: Super-Effizienz  
(Quelle: in Anlehnung an [SCH2000, S. 113])

Dieser Abstand zum effizienten Rand manifestiert sich in einem positiven Super-Effizienzwert und drückt aus, um wie viel Prozent die Inputs erhöht und die Outputs reduziert werden können, bis die Einheit sich wieder auf dem effizienten Rand befindet [SCH2000, S. 112 f.] (Diese minimalen Outputs und maximalen Inputs für  $DMU_2$  lassen sich in der Abbildung anhand der eng gestrichelten Linien ablesen.). Man kann also die Super-Effizienz als ein Maß für die „Übererfüllung“ an Effizienz bezeichnen, die eine DMU im Vergleich zu einer Effizienzbetrachtung der übrigen Einheiten aufweist.

Zur weiteren Veranschaulichung soll anhand derselben Abbildung eine Super-Effizienzbetrachtung der  $DMU_5$  vorgenommen werden: Sie wird in der klassischen Effizienzanalyse als effizient eingestuft und befindet sich dementsprechend auf dem Rand des Technologieraums. Wird sie nun aus der Effizienzbetrachtung genommen, so ergibt sich keine Veränderung der Technologie und ihres Randes.

Somit bleibt der Abstand zu diesem Rand Null, da  $DMU_5$  weiterhin exakt auf dem Rand liegt. Dies entspricht einem Super-Effizienzwert von 0 und macht deutlich, dass diese DMU nicht über eine „Übererfüllung“ von Effizienz verfügt, sondern gerade so effizient war, dass sie eine Position auf dem effizienten Rand erlangte.

Eine weitere sinnvolle Anwendungsmöglichkeit für die Super-Effizienz findet sich im Bereich dynamischer Analysen: Bei der Betrachtung mehrerer Zeitpunkte lassen sich für den Fall solcher DMUs, die in allen Zeitpunkten effizient waren und somit einen Effizienzwert von 1 erlangten, kaum sinnvolle Interpretationen ihrer Entwicklung über die Zeit hinweg anstellen. Erst die Super-Effizienz erlaubt eine Beurteilung der DMUs über den Status der vollständigen relativen Effizienz hinaus und somit auch die weitergehende Beurteilung ihrer Leistung [HOF2006, S. 56]. Inzwischen sind Super-Effizienzmodelle als Basis für ein fundiertes Ranking der effizienten Einheiten einer DEA-Untersuchung allgemein akzeptiert und etabliert. Eine vertiefende Darstellung dieses Themenkomplexes findet sich bei COOPER ET AL. [CST2006, S. 301 ff.].

#### *2.2.5.4 Kreuzeffizienzen und Peer-Orientierung*

In den Abschnitten 2.2.2 und 2.2.3 wurden die Grundmodelle der DEA ausführlich erläutert. Nun soll insbesondere die Frage nach der Gewichtung der In- und Outputs der einzelnen DMUs noch einmal aufgegriffen und kritisch hinterfragt werden: Die bislang eingeführten CCR- und BCC-Modelle gestatten jeder DMU die Anwendung eines für sie individuellen Gewichtesystems – vor allem, um auf diesem Weg eine möglichst hohe Akzeptanz für die Effizienzanalyse zu schaffen. Schließlich hat jede DMU die Möglichkeit, die Gewichte so festzulegen, dass die eigenen Schwächen unter- und die Stärken überbewertet werden, was die Erzielung eines höchstmöglichen Effizienzwertes sicherstellt. RÖDDER UND REUCHER bemängeln jedoch, dass diese eigene Einschätzung nach dem Selbstlob-Prinzip (Self-appraisal) nicht objektiv ist und die Vergleichbarkeit der Effizienzergebnisse der einzelnen DMUs in Frage stellt [ROR2009, S. 2].

In der DEA-Literatur besteht grundsätzlich Einigkeit darüber, dass die bloße Ausrichtung auf Best-practice-DMUs nicht die gesamte Information ausschöpft, so lange nur die Effizienzwerte der einzelnen DMUs auf Basis der selbst gewählten Gewichtungen betrachtet werden. Allerdings wird dieser Aspekt nur in vergleichsweise wenigen Arbeiten aufgegriffen und hinterfragt. Zwar stellten DOYLE UND GREEN 1994 [DOG1994] mit der Kreuzeffizienzmatrix einen Ansatz vor, der die Systematik der individuell optimalen Wahl der Gewichte (Self-appraisal) um weitere Effizienzinformationen ergänzt und somit einen breiteren Blickwinkel auf die Effizienzen der einzelnen DMUs ermöglicht. Bezüglich der konkreten Auswertung und weiterer Einsatzmöglichkeiten dieser Kreuzeffizienzmatrix finden sich jedoch nur einige wenige Veröffentlichungen.

Konkret bestimmt im Rahmen einer Kreuzeffizienzmatrix weiterhin jede  $DMU_k$  die für sie selbst optimalen Gewichte, um so einen maximalen Effizienzwert zu erreichen. Um aber die betrachtete Datenbasis zu erweitern, werden außerdem die Effizienzwerte aller anderen DMUs aus dem betrachteten Set mit genau diesen Gewichten errechnet. In Anlehnung an DOYLE UND GREEN [DOG1994] lässt sich die Kreuzeffizienzmatrix mit den bewerteten DMUs  $j$  (als Spalten) und den bewertenden DMUs  $k$  (als Zeilen) darstellen:

		Bewertete DMU				
		$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
Bewertende DMU	$k = 1$	E <sub>11</sub>	E <sub>12</sub>	E <sub>13</sub>	E <sub>14</sub>	E <sub>15</sub>
	$k = 2$	E <sub>21</sub>	E <sub>22</sub>	E <sub>23</sub>	E <sub>24</sub>	E <sub>25</sub>
	$k = 3$	E <sub>31</sub>	E <sub>32</sub>	E <sub>33</sub>	E <sub>34</sub>	E <sub>35</sub>
	$k = 4$	E <sub>41</sub>	E <sub>42</sub>	E <sub>43</sub>	E <sub>44</sub>	E <sub>45</sub>
	$k = 5$	E <sub>51</sub>	E <sub>52</sub>	E <sub>53</sub>	E <sub>54</sub>	E <sub>55</sub>

Tabelle 2.1: Aufbau einer Kreuzeffizienzmatrix  
(Quelle: in Anlehnung an [DOG1994, S. 569])



$E_{32}$  steht in dieser Kreuzeffizienzmatrix dementsprechend für den Effizienzwert der  $DMU_2$  unter Verwendung der von  $DMU_3$  ausgewählten Gewichte. Die (weiß hervorgehobene) Diagonale in der Matrix ( $E_{11}, E_{22}, \dots, E_{55}$ ) steht für die Effizienzwerte, die jeweils unter Verwendung der selbstbestimmten Gewichte entstanden sind (Self-appraisal) und bereits in den Grundlagenmodellen im Zentrum der Betrachtung standen; man vergleiche dazu das Quotientenproblem (CCR-I-Q) in Abschnitt 2.2.2.

Jedes  $(\mathbf{u}_k^*, \mathbf{v}_k^*)$  bezeichne als optimale Lösung von Aufgabe (CCR-I-Q) ein Gewichtungssystem, mit dem sich die  $DMU\ k$  selbst bestmöglich einschätzt (Self-appraisal). Die nicht grau hinterlegten Felder der Kreuzeffizienzmatrix in Tabelle 2.1 errechnen sich dadurch, dass die Effizienz jeder  $DMU$  unter den Gewichten jeder anderen  $DMU$  des Sets nach demselben DEA-Verfahren ermittelt wird. Das heißt, mit jedem  $(\mathbf{u}_k^*, \mathbf{v}_k^*)$  lassen sich nun auch alle (anderen)  $DMUs\ j$  aus Sicht der  $DMU\ k$  hinsichtlich deren Output-/Inputrelationen beurteilen:

$$h_{kj}^* = \frac{\mathbf{u}_k^{*T} \mathbf{y}_j}{\mathbf{v}_k^{*T} \mathbf{x}_j}$$

Nimmt der Quotient für ein  $j$  den Wert 1 an, so ist  $j$  Referenz- $DMU$  von  $DMU\ k$  – allgemein auch als Best-practice- $DMU$  bezeichnet. Für jede  $DMU_k$  gibt es modellgemäß solche Referenz- $DMUs$ , auf deren Basis die relative Effizienz bestimmt wird. Somit ist in jeder Zeile der Kreuzeffizienzmatrix zwangsläufig mindestens eine 1 zu finden.

Tabelle 2.2 zeigt das von RÖDDER UND REUCHER übernommene Datenbeispiel von fünf  $DMUs$  mit jeweils zwei In- und Outputs – dieses konkrete Beispiel soll der weiteren Veranschaulichung der Kreuzeffizienzen dienen:

DMU <sub>j</sub>	$x_{1j}$	$x_{2j}$	$y_{1j}$	$y_{2j}$
1	2,00	5,00	1,00	2,00
2	2,00	4,00	2,00	3,00
3	8,00	5,00	2,00	2,00
4	3,00	4,00	1,00	2,00
5	6,00	2,00	2,00	1,00

Tabelle 2.2: In- und Outputs von fünf DMUs  
(Quelle: [RRE2011, S. 1262])

Unter Verwendung dieser Daten kommen RÖDDER UND REUCHER [RRE2011] auf Basis eines CCR-Modells zu folgender inputorientierter Kreuzeffizienzmatrix:

		Bewertete DMU				
		$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
Bewertende DMU	$k = 1$	0,67	1,00	0,17	0,44	0,11
	$k = 2$	0,50	1,00	0,60	0,63	1,00
	$k = 3$	0,50	1,00	0,60	0,63	1,00
	$k = 4$	0,53	1,00	0,53	0,67	0,67
	$k = 5$	0,42	1,00	0,56	0,45	1,00

Tabelle 2.3: Kreuzeffizienzmatrix, inputorientiert  
(Quelle: [RRE2011, S. 1262])

Als Problem erweist sich im Zusammenhang mit der Kreuzeffizienzmatrix, dass die Gewichte ( $u_k^*$ ,  $v_k^*$ ) nicht eindeutig sind und sich somit bei mehrfacher Be-

rechnung ceteris paribus unterschiedliche Kreuzeffizienzmatrizen ergeben können. DOYLE UND GREEN tragen dem Rechnung, indem sie die Optimierungsaufgabe (CCR-I-Q) um eine zusätzliche Bedingung erweitern [DOG1994], wobei hier im Allgemeinen auf die Maximierung der Kreuzeffizienzen aller anderen DMUs zurückgegriffen wird.

Die in der Kreuzeffizienzmatrix zur Verfügung stehende breite Informationsbasis zu den Effizienzen der DMUs aus verschiedenen Perspektiven bedarf nun einer transparenten Form der Auswertung, damit eine zielgerichtete Interpretation möglich wird. Schon DOYLE UND GREEN arbeiteten 1994 [DOG1994] mit verschiedenen Formen von Reihungen, um damit die Kreuzeffizienzmatrix mit unterschiedlichen Zielrichtungen aufzubereiten und auf dieser Basis einen gemeinsamen Peer für alle DMUs festzulegen. REUCHER ET AL. stellen einen umfassenden Katalog der verschiedenen Ansätze zur Bildung von Reihungen vor [RRL2008, S. 296 ff.]: Reihungen nach der eigenen, maximalen Selbsteinschätzung haben sich unter dem Begriff „Eigenlob“ etabliert. Am Beispiel der Kreuzeffizienzmatrix in Tabelle 2.3 ergibt sich konkret folgende Reihung:

Rang	1	2	3	4	5
DMU <sub>j</sub>	2	5	1	4	3
eff <sub>j</sub>	1,00	1,00	0,67	0,67	0,60

**Tabelle 2.4: Reihung nach „Eigenlob“**  
(Quelle: in Anlehnung an [RRL2008, S. 297])

Das Prinzip „Alleinstellung“ basiert auf dem sogenannten Maverick-Koeffizienten, „der die relative Überlegenheit der eigenen Effizienz zur eigenen Einschätzung der Konkurrenz ausdrückt“ [RRL2008, S. 297]. Diese Reihung ist also geprägt von der selbst empfundenen Überlegenheit einer DMU gegenüber den übrigen Einheiten, das heißt, es steht derjenige vorne, der den eigenen Vorsprung gegenüber den anderen als besonders groß einschätzt. Diese zeilenorientierte Sicht von REUCHER ET AL. teilen auch APPA ET AL. [AAW2006].

Darüber hinaus steht eine Vielzahl weiterer Reihungsphilosophien zur Verfügung, wie zum Beispiel „Fairness“, „Rücksichtnahme“ oder „Reputation“ [RRL2008, S. 297-300], wobei die Auswahl immer vor dem Hintergrund der jeweiligen Zielrichtung der Untersuchung und deren Ausgestaltung getroffen werden muss. Auch die individuelle Entwicklung einer Reihungsmethode zur Auswahl eines Peers kann in Einzelfällen durchaus Sinn machen. Es bestehen also vielfältige Möglichkeiten, aus der umfangreichen Datengrundlage einer Kreuzeffizienzmatrix mittels Reihungen bzw. Rankings nach unterschiedlichen Aspekten und Schwerpunkten einen Peer zur Bestimmung der relativen Effizienzen aller DMUs zu bestimmen. In der vorliegenden Arbeit sollen die konkreten Reihungen keine weitere Rolle spielen und werden insofern nicht detailliert betrachtet. Gezielt betrachtet werden sollen jedoch die Konsequenzen der Wahl eines Peers, genauer dessen Gewichtungssystems, als Referenz aller DMUs im Vergleich zum (in der Anwendung üblichen) Self-appraisal. Der Fokus soll neben der reinen Effizienzwirkung insbesondere auf der Frage der allokativen Konsequenzen eines Peer-appraisal liegen – eine Fragestellung, die, entgegen der Peer-Bestimmung selbst, in der DEA-Literatur in diesem Zusammenhang kaum betrachtet wird.

RÖDDER UND REUCHER veranschaulichen in ihrem Artikel [RRE2011] die unterschiedlichen Effizienzeinflüsse von Self-appraisal und Peer-appraisal an einem konkreten Zahlenbeispiel, welches in Abbildung 2.30 graphisch dargestellt ist. Diese zeigt ein Szenario mit zwei Inputs und einem Output, auf den zugunsten einer höheren Anschaulichkeit normiert wurde, sowie ein beispielhaft gegebenes Set von DMUs, genauer gesagt deren Aktivitäten, und den Rand der Technologie, auf dem sich fünf Best-practice-DMUs befinden:

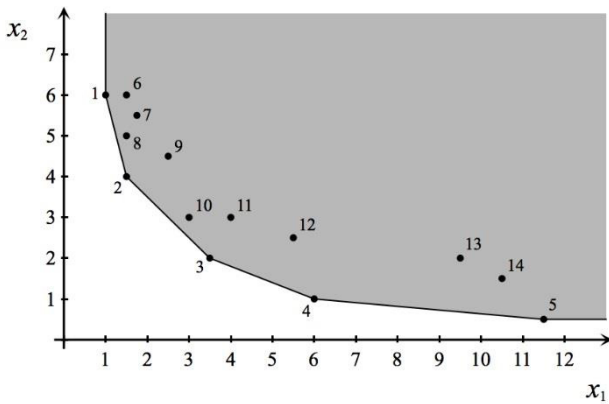


Abbildung 2.30: Aktivitäten eines beispielhaften Sets von DMUs  
(Quelle: in Anlehnung an [ROR2011, S. 150])

Führt man für dieses Szenario eine inputorientierte Optimierung auf der Basis radialer Inputreduktion durch, so ergibt sich die in Abbildung 2.31 dargestellte Vielzahl unterschiedlicher effizienter Zielpunkte für die einzelnen DMUs:

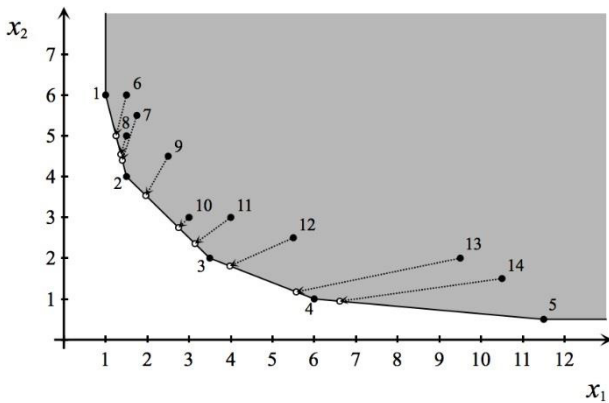


Abbildung 2.31: Effizienzoptimierung im Rahmen des Self-appraisal bei radialer Inputreduktion  
(Quelle: in Anlehnung an [ROR2011, S. 150])

Dabei zeigt sich, dass alle ineffizienten DMUs aufgrund der Vorgabe einer rein radialen Inputreduktion auf den Nullpunkt zulaufen, bis sie auf den effizienten Rand der Technologie treffen und damit effizient sind.

Ein anderes Bild ergibt sich auf Basis der Möglichkeit freier Inputvariation, also einer freien Reduktion der einzelnen Inputs unabhängig voneinander:

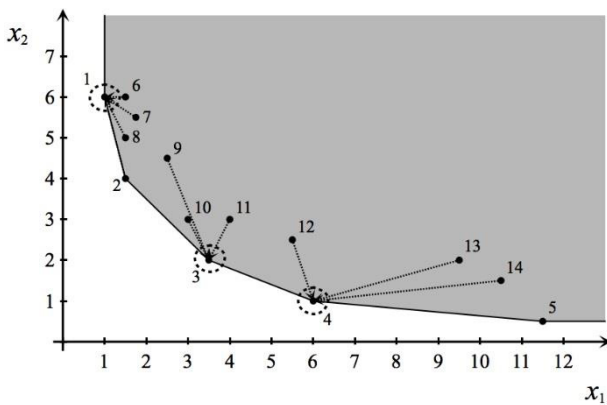


Abbildung 2.32: Clusterbildung im Rahmen des Self-appraisal bei freier Inputvariation  
(Quelle: eigene Darstellung)

Anhand der Berechnungen von RÖDDER UND REUCHER [RRE2011] kommt es zu der in Abbildung 2.32 dargestellten Orientierung der ineffizienten Einheiten, die auf Basis der Self-appraisal-Gewichte immer die ihnen ähnlichen Best-practice-Einheiten auf dem effizienten Rand (im Beispiel die DMUs 1, 3 und 4) als Bewertungsmaßstab heranziehen und sich diesen im Rahmen einer späteren Effizienzverbesserung annähern – man spricht dabei von einer Clusterbildung [ROR2009, S. 8] oder dem Merging-Effekt. Im Rahmen dieser Verlagerung der Aktivitäten aller DMUs in einige wenige Cluster kommt es neben der bereits aufgezeigten Effizienzwirkung auch zu massiven Beeinflussungen der Allokation der Vergleichsgrößen der einzelnen DMUs. Abbildung 2.32 zeigt anschaulich,

dass im Zuge der Effizienzverbesserungen einzelner DMUs auch eine (teilweise extreme) Angleichung des Einsatzverhältnisses der beiden Inputs  $x_1$  und  $x_2$  stattfindet.

Zur Einordnung der Abbildung 2.33 lohnt ein Blick auf Abbildung 2.12 (auf Seite 32), die eine Isokostengerade zeigt, um die technische und allokativen Effizienz herzuleiten. Analog dazu sehen wir hier die bekannten DMUs in der auf den Output normierten Darstellung, ergänzt um verschiedene (nicht eindeutige) Isokostengeraden der Best-practice-DMUs:

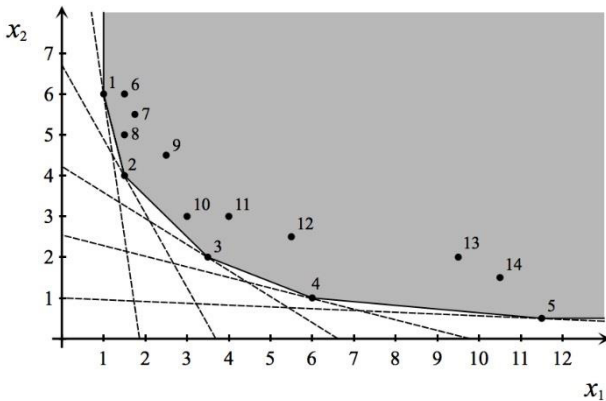


Abbildung 2.33: Isokostengeraden der Best-practice-DMUs  
(Quelle: eigene Darstellung)

Bei der Auswahl eines für alle DMUs verbindlichen Peers sollte neben den bislang angeführten Aspekten auch die Frage eine Rolle spielen, wie gut eine potenzielle Peer-DMU mit den eingesetzten Ressourcen umgeht. Dieser Aspekt der kostenoptimalen Allokation wurde bislang im Rahmen der DEA nicht fokussiert und drängt sich als Ansatzpunkt weiterer Überlegungen auf. Bereits in der Diskussion der Effizienzoptimierung im Rahmen des Self-appraisal und der damit einhergehenden Clusterbildung wurde deutlich, dass die Verbesserung der Effizienzwerte (bei freier Inputvariation) auch deutlichen Einfluss auf die Allokation

der zu optimierenden Vergleichsgrößen hat – das Gewichtungssystem und dementsprechend die Isokostengeraden der jeweiligen Ziel-DMUs werden für die betreffenden ineffizienten DMUs maßgeblich.

Im Fall einer Peer-Orientierung sind grundsätzlich analoge Wirkungszusammenhänge zu beobachten: Bezüglich der Effizienzwirkung tritt derselbe Effekt auf, hat aber für einige DMUs weitaus massivere Allokationswirkungen. Wie RÖDDER UND REUCHER rechnerisch zeigen, ergeben sich im Fall rein radialer Inputreduktion keine Unterschiede zwischen Self-appraisal und Peer-appraisal [ROR2011, S. 151]: Die DMUs laufen auch im Peer-appraisal in einer gedachten Linie zum Ursprung des Koordinatensystems bis auf den effizienten Rand der Technologie, wie es schon in Abbildung 2.31 für den Fall des Self-appraisal gezeigt wurde. Self-appraisal und Peer-appraisal haben hier dieselben Wirkungen bezüglich Effizienz und Allokation.

Im Fall freier Inputvariation reduziert sich durch die Orientierung aller DMUs an einem gemeinsamen Peer die Anzahl der im Rahmen des Self-appraisal eingeführten Cluster (siehe Abbildung 2.32) auf einen gemeinsamen Zielpunkt. Alle DMUs müssen sich dem Gewichtungssystem des vorher bestimmten Peers „unterwerfen“ und dessen Isokostengerade(n) wird/werden für alle DMUs relevant. Man spricht dabei von Konsensualer Effizienz, die in Abbildung 2.34 graphisch dargestellt ist.

Da sich letztendlich alle DMUs auf die Aktivität der Peer-DMU bewegen, liegt hier ein maximales Merging vor: Nach abgeschlossener Effizienzverbesserung haben alle Aktivitäten der einzelnen DMUs dieselben Koordinaten, in Abbildung 2.34 die der DMU<sub>3</sub>. Diese vereint zwei Eigenschaften in sich, nämlich ihre Lage auf dem effizienten Rand sowie eine kostenoptimale Allokation der Ressourcen  $x_1$  und  $x_2$ , gemäß der eingezeichneten Isokostengeraden.



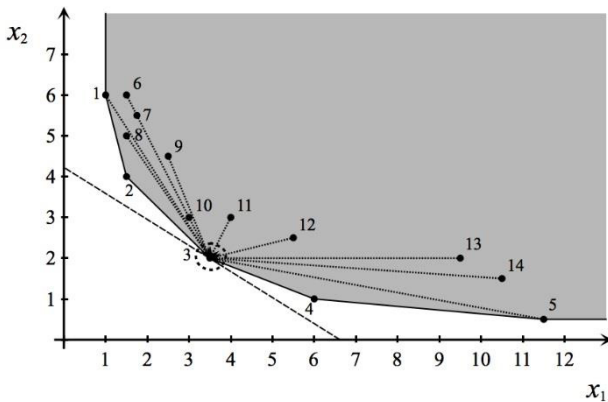


Abbildung 2.34: Effizienzverbesserungen unter Peer-appraisal und freier Inputvariation  
(Quelle: eigene Darstellung)

WU ET AL. [WSL2011, S. 5162] sehen in der Verwendung von Kreuzeffizienzen neben der Peer-Orientierung und den daraus folgenden Ranking-Möglichkeiten weitere Vorteile, insbesondere die Identifikation extremer Gewichtungssysteme und die Differenzierungsmöglichkeiten zwischen den als effizient klassifizierten Einheiten. Zusammenfassend kann man die Kreuzeffizienzen als eine in mehrfacher Hinsicht sinnvolle Weiterentwicklung der klassischen DEA-Modelle bezeichnen, da sie einen breiteren Blick auf die Untersuchungsergebnisse liefern und zur Objektivierung der Ergebnisinterpretation beitragen können. Somit können Kreuzeffizienzen auch ein wichtiges Mittel sein, um die notwendige Akzeptanz bei den Teilnehmern von DEA-Untersuchungen herzustellen bzw. zu steigern.

## 2.2.6 Reflektion der DEA als Instrument der Effizienzmessung

In Abschnitt 2.1.2 wurde eine erste Systematisierung der unterschiedlichen Methoden der Effizienzmessung vorgenommen, bei der insbesondere die Kategorisierung nach parametrischen bzw. nicht-parametrischen und deterministischen

bzw. stochastischen Ansätzen im Vordergrund stand (siehe dazu auch Abbildung 2.18 auf Seite 41). Nun soll speziell die DEA in den Fokus gestellt und dabei auch in Relation zu alternativen Verfahren gesetzt werden.

Eines der Hauptziele dieser Arbeit liegt in der Fundierung einer situativ angemessenen, bewussten und zielgerichteten Anwendung der etablierten DEA-Modelle. Die nachfolgenden Ausführungen sollen vor diesem Hintergrund auch und insbesondere als Grundlage einer jeden DEA-Untersuchung dienen, nämlich bei der einleitenden Fragestellung, ob die DEA in der individuellen Ausgangslage und Zielsetzung überhaupt das geeignete Instrument ist und den Erwartungen der unterschiedlichen Anspruchssteller gerecht werden kann. Dazu werden die allgemeinen Stärken und Schwächen der DEA im Vergleich zu alternativen Methoden der Effizienzmessung herausgearbeitet, bevor in Kapitel 4 die konkreten Anforderungen und Ausgestaltungsmöglichkeiten einer DEA-Untersuchung diskutiert werden.

Neben der DEA ist insbesondere die Stochastic Frontier Analysis (SFA) von empirischer Relevanz [CRO2005, S. 312], weshalb diese hier der DEA gegenübergestellt und in ihren prägenden Eigenschaften dargestellt werden soll. Bereits im Unterabschnitt 2.1.2.2 wurde die SFA als parametrisches Verfahren der Effizienzmessung eingeführt, was schon den grundlegenden und wichtigsten Unterschied zur DEA ausmacht: Als parametrisches Verfahren ist die SFA darauf angewiesen, dass die Produktionsfunktion als Grundlage der Effizienzberechnung gegeben ist. Die exakte Formulierung einer Produktionsfunktion für reale Beobachtungseinheiten mit mehreren In- und Outputs dürfte in der Praxis kompliziert bis unmöglich sein [SCH2006, S. 95], weshalb sich eine sinnvolle Anwendung der SFA auf spezielle Randbereiche mit geeigneten Bedingungen reduzieren dürfte. Insofern ist die (inhärente) Approximation der Produktionsfunktion im Rahmen der DEA als Vorzug zu bewerten, der eine Anwendung bei unbekannter Produktionsfunktion ermöglicht [HAA2002, S. 45] und die vorgelagerte Unterstellung jeglicher Annahmen zur Produktionsfunktion obsolet macht [BCC1985]. Tabelle 2.5 gibt einen Überblick über die Eignung beider Verfahren

in Bezug auf unterschiedlichste Kriterien. Dazu und zu den begleitenden Erläuterungen sei, neben den zitierten Stellen, auch auf die Arbeit von SCHLAMP [SCH2006] verwiesen.

	DEA	SFA
Berücksichtigung struktureller Unterschiede im Set	↑	↗
Aufnahme unterschiedlich skalierten Daten	↑	↑
Berücksichtigung unkontrollierbarer Variabler	↑	↓
Objektive und individuelle Gewichtung der In- und Outputfaktoren	↑	↘
Akzeptanz bei den Anspruchstellern	↑	→
Bestimmung von Best-practice-Unternehmen	↑	↑
Ableitung konkreter Handlungsmöglichkeiten	↗	↘
Wirtschaftlichkeit der Ergebniserstellung	↗	→
Robustheit gegenüber statistischen Fehlern	↓	↑

Legende: ↑ = sehr gut, ↗ = gut, → = mittelmäßig, ↘ = schlecht, ↓ = sehr schlecht

**Tabelle 2.5: Kriterienbezogener Vergleich von DEA und SFA**  
(Quelle: in Anlehnung an [SCH2006, S. 96])

Sowohl die DEA als auch die SFA bestimmen den effizienten Rand nicht über eine einzelne Betrachtungseinheit, so dass strukturelle Unterschiede in die Betrachtung mit einfließen. Die DEA geht hier allerdings noch ein Stück weiter, da die einzelnen Einheiten mit denjenigen Einheiten verglichen werden, die (graphisch) in ihrer Nähe liegen, ihnen also ähnlich sind. Ein großer Vorteil der DEA liegt in ihrer Fähigkeit, mehrere In- und Outputs in einer Untersuchung zu berücksichtigen, die in unterschiedlichen Einheiten gemessen bzw. angegeben werden. So können beispielsweise Beträge in einer Währung, Mitarbeiter nach Kopfzahl und die Erfüllung qualitativer Ziele problemlos in dieselbe DEA-

Untersuchung einfließen. Außerdem erlaubt die DEA die Berücksichtigung unkontrollierbarer Variabler, wie in Unterabschnitt 2.2.5.2 aufgezeigt wurde. Eine objektive Gewichtung der In- und Outputfaktoren wird bei der DEA in besonderem Maß gewährleistet: Erstens werden für jede betrachtete DMU individuelle Gewichte ermittelt und zweitens ergeben sich diese im Rahmen der Optimierung, sind also keinerlei subjektiven Einflüssen ausgesetzt. Die SFA hingegen legt an alle DMUs dieselben Gewichte an [BIE2009, S. 9]. Als Verfahren der relativen Effizienzmessung genießt die DEA eine vergleichsweise hohe Akzeptanz bei den Teilnehmern der Analyse, da jede DMU an Konkurrenten gemessen wird, statt an einer fiktiven und eventuell unrealistisch effizienten Einheit. Dies gilt auch für die SFA, soweit alle Beteiligten die Vorgabe der Produktionsfunktion im konkreten Untersuchungsszenario als realitätsnah einschätzen. Auch akzeptanzfördernd ist die Tatsache, dass sich aus den Ergebnissen einer DEA konkrete und individuelle Maßnahmen ableiten lassen, nämlich inwieweit Inputs reduziert und/oder Outputs erweitert werden müssen. Eine Bestimmung/Identifizierung der Best-practice-DMUs gewährleisten beide Verfahren. Im Rahmen einer DEA-Untersuchung kann der Vergleich mit einem Konkurrenten bei einer statischen Betrachtung jedoch eine gewisse Ungerechtigkeit mit sich bringen, da dieser nichts über die (zum Zeitpunkt der Datenerhebung) maximal erreichbare Effizienz der betrachteten DMU aussagt [HAA2002, S. 46]. Eventuell muss diese DMU erst durch Anpassung ihrer Prozesse und Strukturen die eigene Transformation so beeinflussen, dass die Effizienz der Best-practice-DMU erreichbar wird. Diese Tatsache sollte jedoch nicht als Nachteil gewertet werden, da die DEA auch das Ziel verfolgt, den DMUs Potenziale aufzuzeigen, die dann in einem dynamischen Prozess gehoben werden können. Oft wird der rechnerische Aufwand als Schwäche der DEA genannt, da für jede DMU ein separates Optimierungsproblem gelöst werden muss. In Zeiten immer leistungsfähigerer Prozessoren scheint dieser Kritikpunkt von geringer Relevanz. Von größerer Bedeutung ist die starke Anfälligkeit der DEA gegenüber Datenfehlern [RAD2008, S. 51] [PHP2004, S. 19] [CRO2005, S. 313]. Diese sind im Rahmen der DEA vor allem deshalb so gefährlich, weil sie ggf. (unrechtmäßige) Best-practice-Einheiten generieren und darüber die gesamten Untersuchungsergeb-

nisse massiv verfälschen können. Auch deshalb muss bei Auswahl, konkreter Definition und Erhebung der notwendigen Daten besondere Sorgfalt herrschen. An dieser Stelle kann die SFA punkten, die in der Lage ist, Ineffizienzen von statistischen Fehlern zu trennen [LOS1988, S. 23].

Insgesamt darf die DEA als fest etabliertes Instrument der relativen Effizienzmessung betrachtet werden, das inzwischen in den unterschiedlichsten Bereichen Anwendung findet. Das nächste Kapitel gibt einen Überblick über die empirischen Anwendungen der DEA und analysiert eine Zahl solcher Arbeiten im Hinblick auf verschiedene Gesichtspunkte.

Praxisorientierte Data Envelopment Analysis

Kerpen, P.

2016, XV, 331 S. 85 Abb., Softcover

ISBN: 978-3-658-15459-2