2 Grundlagen der Effizienzmessung und die Data Envelopment Analysis

2.1 Der Begriff der Produktion und Grundbegriffe der Aktivitätsanalyse

Die *Produktion* bezeichnet einen Prozess der betrieblichen Leistungserstellung, bei welchem durch die Kombination bzw. Umwandlung von Gütern neue Güter erzeugt werden.[1] Die in die Produktion eingehenden Güter werden auch als Produktionsfaktoren oder Inputgüter, kurz Inputs, bezeichnet, die erzeugten Güter werden als Produkte oder Outputgüter, kurz Outputs, bezeichnet. Für den Produktionsprozess werden auch die Begriffe Faktorkombinationsprozess oder Throughput verwendet. In der Produktionstheorie werden Produktionssysteme oft wie in Abbildung 2.1 als so genannte Input-Throughput-Output-Systeme dargestellt.[2]

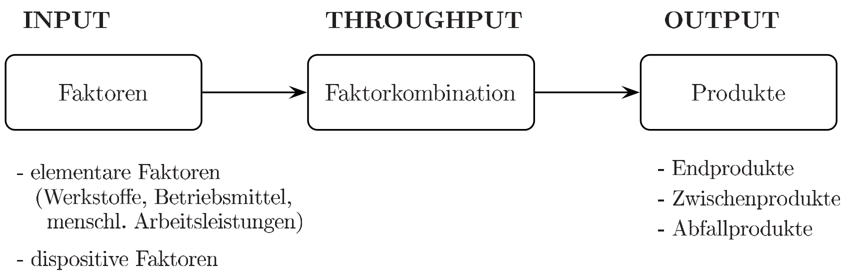

Abb. 2.1: Input-Throughput-Output-System

Der *Input* eines Produktionssystems sind die eingesetzten Güter bzw. die Produk-

[1] Vgl. u. a. WITTMANN 1968, S. 2; FANDEL 2007, S. 5.

[2] Zur Darstellung des Input-Throughput-Output-Systems einer Produktion vgl. u. a. STEVEN 1998, S. 3; DINKELBACH/ROSENBERG 2004, S. 2; FANDEL 2007, S. 1.

R. Schlindwein, *Effizienzanalyse von Dienstleistungsproduktionen*,
DOI 10.1007/978-3-658-14322-0_2

tionsfaktoren. In der Betriebswirtschaftslehre gibt es zahlreiche Versuche, die Produktionsfaktoren zu systematisieren, wobei üblicherweise vom Produktionsfaktorsystem Gutenbergs ausgegangen wird.[3] Bei diesem wird auf der obersten Ebene eine Einteilung in elementare und dispositive Produktionsfaktoren vorgenommen. Die Absicht dabei ist, zwischen den Faktoren zu unterscheiden, die in die Produktion einfließen (elementare Faktoren), und denjenigen Faktoren, die den Produktionsprozess gestalten (dispositive Faktoren). In Anlehnung an Gutenberg können die elementaren Produktionsfaktoren wiederum in drei Gruppen eingeteilt werden:[4] Werkstoffe werden direkt eingesetzt, d. h. sie werden zu einem Bestandteil der Produkte.[5] Betriebsmittel (z. B. Maschinen) dienen indirekt der Produktion, indem sie Verrichtungen abgeben. Die dritte Gruppe sind menschlichen Arbeitsleistungen, die produktionsbezogen, aber nicht dispositiv sind.

Der *Throughput* bezeichnet einen Prozess, bei welchem die Produktionsfaktoren (Inputs) in die Produkte (Outputs) überführt werden.[6] Dazu werden die Produktionsfaktoren kombiniert und in neue Güter transformiert. Dieser Transformationsprozess erfolgt nach bestimmten technischen Gesetzmäßigkeiten, die physikalischer, chemischer, biologischer oder anderer Art sein können.[7]

Der *Output* des Produktionssystems sind die erzeugten Güter bzw. die Produkte.[8] Man unterscheidet Endprodukte, Zwischenprodukte und Abfallprodukte.[9] Endprodukte werden nach der Herstellung vom Unternehmen an andere Wirtschaftssubjekte abgegeben und lassen sich in Konsumgüter und Investitionsgüter gliedern. Zwischenprodukte werden bei mehrstufiger Fertigung im Unternehmen als Produktionsfaktoren weiterverwendet. Abfallprodukte entstehen bei der Güterherstellung oder -verwertung, werden aber nicht mehr zum Konsum oder zur Produktion genutzt. Die Erbringung des Output in Form von neuen Produkten (Endprodukten) ist der Zweck des Produktionsprozesses und dient letztendlich der Nutzenstiftung bei den Kunden.[10]

Die Beschreibung der quantitativen Beziehungen zwischen den Produktionsfak-

[3] Vgl. auch im Weiteren GUTENBERG 1983, S. 11ff; CORSTEN 2007, S. 4f.

[4] Zu den elementaren Produktionsfaktoren vgl. u. a. GUTENBERG 1983, S. 2; STEVEN 1998, S. 3f; FANDEL 2007, S. 34. Zu einer Untergliederung der Elementarfaktoren entsprechend ihres Beitrags zur Leistungserstellung in Verbrauchsfaktoren und Potenzialfaktoren (auch Gebrauchs-, Bestandsfaktoren) vgl. u. a. FANDEL 2007, S. 34.

[5] Zur Einteilung der Werkstoffe in Roh-, Hilfs- und Betriebsstoffe vgl. u. a. STEVEN 1998, S. 3.

[6] Vgl. u. a. CORSTEN 2007, S. 8.

[7] Vgl. FANDEL 2007, S. 1.

[8] Vgl. STEVEN 1998, S. 1; FANDEL 2007, S. 1.

[9] Vgl. auch im Weiteren WITTMANN 1968, S. 2; FANDEL 2007, S. 33f. Eine andere Einteilung der Produkte ist z. B. möglich, indem er Hauptprodukte (Sachleistungen, Dienstleistungen) von Nebenprodukten unterschieden werden. Vgl. DYCKHOFF 2006, S. 46.

[10] Vgl. DYCKHOFF et al. 2007, S. 6.

toren (Inputs) und den Produkten (Outputs) ist die Aufgabe der *Produktionstheorie*.[11] Dazu werden Produktionsmodelle formuliert, welche diese Beziehungen mit Hilfe von Technologien oder den hieraus abgeleiteten Produktionsfunktionen oder -korrespondenzen abbilden.[12] Eine Möglichkeit zur Herleitung solcher Aussagenzusammenhänge stellt die *Aktivitätsanalyse* dar.[13] Sie geht auf KOOPMANS 1951, DEBREU 1965 und HILDENBRAND 1966 zurück und ermöglicht eine formal präzise Darstellung von Input-Output-Beziehungen mit Hilfe von Technologien, wie im Folgenden gezeigt wird.[14]

Bezeichnet man x_m als die Menge des m-ten Produktionsfaktors (Inputs), die bei einer Produktion eingesetzt wird,[15] dann kann die Menge der M Produktionsfaktoren als Vektor dargestellt werden:

$$\mathbf{x} := \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} \in \mathbb{R}_+^M ,$$

wobei $\mathbb{R}_+$ die Menge der nichtnegativen reellen Zahlen ist.

Bezeichnet man y_n als die Menge des n-ten Produktes (Outputs), die bei einer Produktion erzeugt wird,[16] dann ist der Vektor der N Produkte:

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}_+^N .$$

Der Zusammenhang zwischen den Input- und Outputmengen (Faktoreinsatz- und Ausbringungsmengen) bei einer möglichen Produktion wird durch die *Aktivität* $\mathbf{z}$ erfasst. Im $(M+N)$-dimensionalen Raum wird eine Aktivität durch einen Punkt repräsentiert, weshalb sie auch als Produktionspunkt oder kurz Produktion bezeichnet wird. Da Inputs in die Produktion eingehen, Outputs aus der Produktion hervorgehen, bedürfen sie einer unterschiedlichen Handhabung, welche durch

[11] Vgl. STEVEN 1998, S. 5; FANDEL 2007, S.32.
[12] Vgl. CORSTEN 2007, S. 48; FANDEL 2007, S.32.
[13] Vgl. auch im Weiteren SCHWEITZER/KÜPPER 1997, S. 41.
[14] Zu den folgenden Grundbegriffen der Aktivitätsanalyse vgl. u. a. WITTMANN 1968, S. 2f; TURETSCHEK 1981, S. 68ff; FANDEL 1990, S. 4f; JAHNKE 1995, S. 49f (insbesondere zur Grundaktivität); SCHWEITZER/KÜPPER 1997, S. 41; KLEINE 2002, S. 69ff; CORSTEN 2007, S. 60ff; FANDEL 2007, S. 35ff.
[15] Im Weiteren auch als (Faktor-)Einsatzmenge, Inputmenge oder Inputquantität bezeichnet.
[16] Im Weiteren auch als Ausbringungsmenge, Outputmenge oder Outputquantität bezeichnet.

die Verwendung unterschiedlicher Vorzeichen erfolgt:

$$\mathbf{z} := \begin{pmatrix} -x_1 \\ \vdots \\ -x_M \\ y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} -\mathbf{x} \\ \mathbf{y} \end{pmatrix} \in -\mathbb{R}_+^M \times \mathbb{R}_+^N =: \mathbb{R}_{-+}^{M+N}.$$

Die Aktivitätsanalyse baut auf so genannten *Grundaktivitäten* auf. Jede Grundaktivitäten beschreibt eine elementare Verfahrensweise eines Unternehmens. In der linearen Aktivitätsanalyse wird angenommen, dass die Grundaktivitäten mit verschiedenen Intensitäten betrieben werden können. Somit stehen diese Grundaktivitäten stellvertretend für eine Vielzahl von Aktivitäten.

Die Menge aller technisch möglichen Aktivitäten eines Produktionssystems ist die *Technologiemenge* TM oder kurz Technologie:[17]

$$TM := \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \,\middle|\, \mathbf{z} \text{ ist technisch möglich} \right\}.$$

Die Technologie beruht auf technischen Gesetzmäßigkeiten und schränkt die Produktionsmöglichkeiten ein. In der Literatur werden Grundannahmen genannt, die auf ökonomischen Plausibilitätsüberlegungen basieren und die für alle Technologien gelten sollten.[18] Darüber hinaus können Technologien durch bestimmte Eigenschaften unterschieden werden, die zu speziellen Technologieformen führen. Diese Grundannahmen sowie die speziellen Eigenschaften von Technologien werden im folgenden Abschnitt erläutert.

[17] Vgl. u. a. WITTMANN 1968, S. 2f; STEVEN 1998 S. 63; CORSTEN 2007, S.62; HACKMAN 2008, S. 1.

[18] Vgl. STEVEN 1998, S. 63; CORSTEN 2007, S. 62f; FANDEL 2007, S. 38f.

2.2 Die Technologiemenge

2.2.1 Grundannahmen zu den Technologiemengen

In der Literatur werden folgende Grundannahmen genannt, die allgemein für Technologien gelten sollten:[19]

- *Abgeschlossenheit*

  Die Technologiemenge TM ist eine abgeschlossene Menge in $\mathbb{R}^{M+N}$, d. h. der Rand ist Teil der Technologiemenge.

- *Möglichkeit der Untätigkeit*

  Die Untätigkeit bzw. der Produktionsstillstand, gekennzeichnet durch den Nullvektor, ist Element der Technologiemenge:

$$\mathbf{0} \in TM \qquad \text{mit} \quad \mathbf{0} := \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{M+N}.$$

  Der Produktionsstillstand stellt zwar eine technisch mögliche Produktion dar, er kann jedoch durch Restriktionen, z. B. Lieferverpflichtungen, faktisch ausgeschlossen werden.[20]

- *Unmöglichkeit des Schlaraffenlandes*

  Es gibt keine positive Ausbringungsmenge ohne den Einsatz von Produktionsfaktoren:

$$\begin{pmatrix} \mathbf{0} \\ \mathbf{y} \end{pmatrix} \notin TM \qquad \text{für} \quad \mathbf{y} \neq \mathbf{0}.$$

- *Möglichkeit ertragreicher Produktion*

  Eine Technologiemenge soll mindestens eine technisch zulässige Produktion mit einer positiven Ausbringungsmenge enthalten:

$$\exists \mathbf{z} \in TM : y_n > 0 \text{ für mindestens ein } n,\ n = 1, ..., N.$$

[19] Zu den Grundannahmen vgl. u. a. Wittmann 1968, S. 4ff; Schweitzer/Küpper 1997, S. 43f; Steven 1998, S. 63; Dyckhoff 2006, S. 132ff; Corsten 2007, S. 62f; Fandel 2007, S. 38f.

[20] Vgl. Dyckhoff 2006, S. 134.

Diese Forderung soll Technologien ausschließen, die lediglich aus der Untätigkeit bestehen, da eine Betrachtung solcher Technologien aus ökonomischen Gründen nicht sinnvoll ist.

- *Unumkehrbarkeit der Produktion*[21]

  Technisch mögliche Produktionen sind nicht umkehrbar. Der Produktionsstillstand stellt diesbezüglich eine Ausnahme dar. Somit ist die Schnittmenge der Technologie mit ihrer Umkehrung die Untätigkeit:

  $$TM \cap -TM = \{\mathbf{0}\} \quad \text{mit} \quad -TM := \{-\mathbf{z} | \mathbf{z} \in TM\}.$$

  Diese Annahme soll ausdrücken, dass es nicht möglich ist, ein erzeugtes Endprodukt nach Abschluss der Produktion wieder in seine Einsatzgüter zurückzuführen.

2.2.2 *Spezielle Eigenschaften von Technologiemengen*

Es lassen sich Technologien aufzeigen, welche die vorgestellten Grundannahmen erfüllen. Des Weiteren können verschiedene Formen von Technologien aufgrund ihrer Eigenschaften unterschieden werden. Insbesondere folgende Eigenschaften sind von Bedeutung:

Skalen- bzw. Größenvariation

Technologien können durch die Möglichkeit der *Skalen-* bzw. *Größenvariation* unterschieden werden, d. h. durch die möglichen Veränderungen des Skalenniveaus bzw. der Größe einer Produktion $\mathbf{z}$ innerhalb der Technologie.[22] Es sei der Skalar $\lambda \in \mathbb{R}_+$ das Maß für das Skalenniveau, auch als Skalar- oder Linearfaktor bezeichnet. Dann lassen sich folgende drei Eigenschaften von Technologien unterscheiden:

- Die Eigenschaft der *Größendegression* liegt vor, wenn proportionale Verringerungen (Niveausenkungen) einer zulässigen Produktion wieder eine zulässige Produktion ergeben:

  $$\mathbf{z} \in TM, \quad 0 \leqq \lambda \leqq 1 \quad \Rightarrow \quad \lambda \cdot \mathbf{z} \in TM.$$

[21] Auch als „Irreversibilität der Produktion“ bezeichnet.

[22] Zur Skalen- bzw. Größenvariation vgl. u. a. WITTMANN 1968, S. 7f; SCHWEITZER/KÜPPER 1997, S. 44f; DYCKHOFF 2006, S. 59f; FANDEL 2007, S. 40f.

- Die Eigenschaft der *Größenprogression* liegt vor, wenn proportionale Erhöhungen (Niveauerhöhungen) einer zulässigen Produktion wieder eine zulässige Produktion ergeben:

  $$\mathbf{z} \in TM, \quad \lambda \geqq 1 \quad \Rightarrow \quad \lambda \cdot \mathbf{z} \in TM.$$

- Die Eigenschaft der *Größenproportionalität* liegt vor, wenn Niveauänderungen einer zulässigen Produktion sowohl in Form von proportionalen Verringerungen als auch von proportionalen Erhöhungen eine zulässige Produktion ergeben:

  $$\mathbf{z} \in TM, \quad \lambda \geqq 0 \quad \Rightarrow \quad \lambda \cdot \mathbf{z} \in TM.$$

  Somit umfasst die Größenproportionalität sowohl die Größendegression als auch die Größenprogression.

Diese Eigenschaften führen zu den grundlegenden Formen der *größenprogressiven, größendegressiven* und *größenproportionalen* Technologien. Im Zusammenhang mit den später vorgestellten DEA-Technologiemengen werden diese Eigenschaften durch die Begriffe *nicht zunehmende (non-increasing), nicht abnehmende (non-decreasing)* sowie *konstante Skalenerträge (constant returns to scale)* beschrieben.[23]

Additivität

Eine Technologie besitzt die Eigenschaft der *Additivität*, wenn zwei zulässige Produktionen auch gemeinsam ausgeführt werden können, ohne sich gegenseitig zu beeinflussen, sodass ihre Summe auch eine zulässige Produktion darstellt.[24] Dies ist der Fall, wenn verschiedene Produktionen unabhängig voneinander ausgeführt werden können und/oder ein und dieselbe Produktion beliebig häufig ausgeführt werden kann:

$$\mathbf{z}_1, \mathbf{z}_2 \in TM \quad \Rightarrow \quad \mathbf{z}_1 + \mathbf{z}_2 \in TM.$$

Linearität und Konvexität

Linearität liegt vor, wenn eine Technologie sowohl durch Größenproportionalität als auch durch Additivität der Produktionen charakterisiert werden kann.[25] Somit

[23] Vgl. u. a. DYCKHOFF 2006, S. 60.

[24] Zur Additivität vgl. u. a. WITTMANN 1968, S. 11f; TURETSCHEK 1981, S. 73; STEVEN 1998, S. 68; KLEINE 2002, S. 73; DYCKHOFF 2006, S. 61; CORSTEN 2007, S. 63; FANDEL 2007, S. 43.

[25] Zur Linearität vgl. u. a. WITTMANN 1968, S. 13; STEVEN 1998, S. 71ff; DYCKHOFF 2006, S. 61ff; CORSTEN 2007, S. 63; FANDEL 2007, S. 43f.

gehören alle nichtnegativen Linearkombinationen der zulässigen Produktionen zur Technologiemenge, sodass gilt

$$\mathbf{z}_1,\mathbf{z}_2 \in TM, \quad \lambda_1,\lambda_2 \geqq 0 \quad \Rightarrow \quad \lambda_1\mathbf{z}_1+\lambda_2\mathbf{z}_2 \in TM.$$

Mit den dargelegten Grundannahmen und den zusätzlichen Eigenschaften der Größenproportionalität und der Additivität ist die *lineare Technologie* TM_L eindeutig definiert. Ausgehend von den Grundaktivitäten $\mathbf{z}_j$, $j=1,...,J$, lässt sich die lineare Technologie formal wie folgt beschreiben:[26]

$$TM_L := \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \,\middle|\, \mathbf{z} = \sum_{j=1}^{J} \lambda_j \cdot \mathbf{z}_j, \quad \lambda_j \geqq 0, j=1,..,J \right\}.$$

Da diese Technologie eine Kegelform besitzt, mit dem Scheitel im Ursprung, wird sie auch als Kegeltechnologie bezeichnet.[27]

Die lineare Technologie stellt einen Spezialfall der *konvexen Technologien* dar.[28] Eine Technologie ist *konvex*, wenn die Summe der Skalarfaktoren eins ergibt:

$$\mathbf{z}_1,\mathbf{z}_2 \in TM, \quad \lambda_1+\lambda_2=1, \quad \lambda_1,\lambda_2 \geqq 0 \quad \Rightarrow \quad \lambda_1\mathbf{z}_1+\lambda_2\mathbf{z}_2 \in TM.$$

Konvexe Technologien setzen stets die beliebige Teilbarkeit der Aktivitäten voraus.

Im Hinblick auf die noch einzuführenden DEA-Technologiemengen wird die lineare Technologie TM_L auch als Technologie mit *konstanten Skalenerträgen (constant returns to scale, TM^{CRS})* bezeichnet. Eine konvexe Technologie, welche die Konvexkombinationen, nicht aber die Linearkombinationen der Grundaktivitäten enthält, wird auch als Technologie mit *variablen Skalenerträgen (variable returns to scale, TM^{VRS})* bezeichnet.[29]

Verschwendbarkeit

Eine Technologiemenge sollte alle Produktionsmöglichkeiten eines Produktionssystems - im Sinne technisch möglicher Transformationen von Inputs in Outputs - umfassen.[30] Somit führt auch die Güterverschwendung zu technisch zulässigen Produktionen. Unterschieden wird zwischen der Verschwendbarkeit der Inputs und

[26] Vgl. KLEINE 2002, S. 73.
[27] Vgl. u. a. STEVEN 1998, S. 73; KLEINE 2002, S. 73.
[28] Zu konvexen Technologien vgl. u. a. WITTMANN 1968, S. 12; DYCKHOFF 2006, S. 63.
[29] Zu den Technologien TM^{CRS} und TM^{VRS} vgl. Abschnitt 2.5.2, S. 33.
[30] Vgl. FANDEL 2007, S. 38.

der Verschwendbarkeit der Outputs:[31]

- Mit der Eigenschaft der *freien Verschwendbarkeit der Inputs* wird die Möglichkeit beschrieben, bei gleich bleibender Ausbringungsmenge den Faktoreinsatz einzelner oder aller Faktoren zu erhöhen:

$$\begin{pmatrix} -\mathbf{x} \\ \mathbf{y} \end{pmatrix} \in TM, \quad \bar{\mathbf{x}} \geqq \mathbf{x} \quad \Rightarrow \quad \begin{pmatrix} -\bar{\mathbf{x}} \\ \mathbf{y} \end{pmatrix} \in TM.$$

- Mit der Eigenschaft der *freien Verschwendbarkeit der Outputs* wird die Möglichkeit beschrieben, bei gleich bleibendem Faktoreinsatz die Ausbringungsmenge einzelner oder aller Produkte zu reduzieren.

$$\begin{pmatrix} -\mathbf{x} \\ \mathbf{y} \end{pmatrix} \in TM, \quad \bar{\mathbf{y}} \leqq \mathbf{y} \quad \Rightarrow \quad \begin{pmatrix} -\mathbf{x} \\ \bar{\mathbf{y}} \end{pmatrix} \in TM.$$

Für freie Verschwendbarkeit im Allgemeinen lässt sich schreiben:

$$\mathbf{z} \in TM, \quad \bar{\mathbf{z}} \leqq \mathbf{z} \quad \Rightarrow \quad \bar{\mathbf{z}} \in TM.$$

Zwar führt die Annahme der Verschwendbarkeit zu technisch möglichen Produktionen, in unbegrenztem Ausmaß ist sie jedoch sowohl in ökonomischer als auch ökologischer Hinsicht fragwürdig, weshalb gegebenenfalls die Einführung von Restriktionen sinnvoll ist.[32] Wie im Weiteren ersichtlich wird, erweisen sich diese Produktionen zudem als nicht effizient.

Erfüllt eine lineare Technologie die Annahme der freien Verschwendbarkeit der Inputs und Outputs, so ergibt sich die Technologiemenge

$$TM_L^V := \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \,\middle|\, \mathbf{z} \leqq \sum_{j=1}^{J} \lambda_j \cdot \mathbf{z}_j, \quad \lambda_j \geqq 0, j = 1,..,J \right\}.$$

Die Eigenschaften Größenproportionalität, Additivität, Linearität und Verschwendbarkeit werden in den Abbildungen 2.2-2.5 anhand eines Produktionssystems mit einem Input x und einem Output y veranschaulicht.[33] Ausgangspunkt sind jeweils die beiden Grundaktivitäten A und B.

[31] Zur Verschwendbarkeit vgl. u. a. BANKER et al. 1984, S. 1081; SCHEEL 2000, S. 45; DARAIO/SIMAR 2007, S. 21.

[32] Vgl. KLEINE 2002, S. 143.

[33] Zur Darstellung der Größenproportionalität, Additivität, Linearität bei einer Technologie mit zwei Inputs und einem konstanten Output vgl. CORSTEN 2007, S. 64.

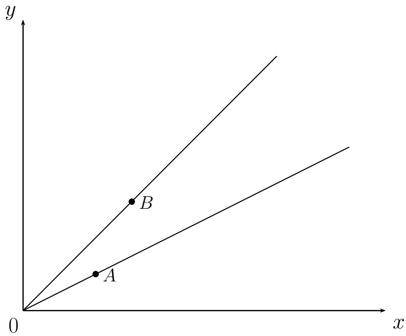

Abb. 2.2: Größenproportionalität

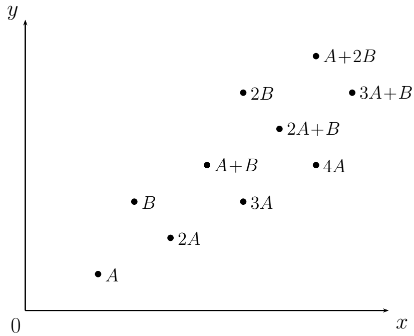

Abb. 2.3: Additivität

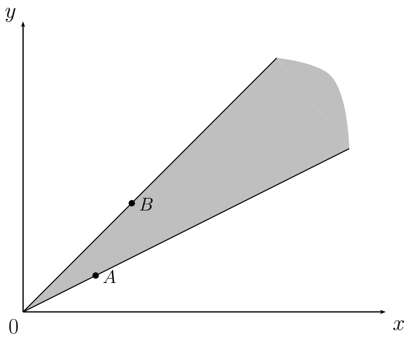

Abb. 2.4: Linearität

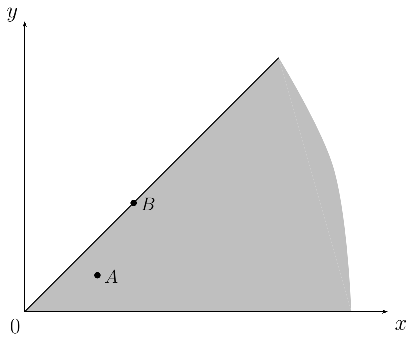

Abb. 2.5: Linearität und Verschwendbarkeit

Restriktionen

Prinzipiell wird in der Aktivitätstheorie angenommen, dass die Technologien die vorgestellten Eigenschaften erfüllen. In der Realität treten jedoch häufig Güterbeschränkungen (*Restriktionen*) auf, sodass nicht alle technisch möglichen Produktionen realisierbar sind.[34] Solche Güterbeschränkungen können zum einen für Produktionsfaktoren bestehen, die im betrachteten Zeitraum in begrenzter Men-

[34] Zur Berücksichtigung von Güterbeschränkungen vgl. u. a. WITTMANN 1968, S. 114ff; TURETSCHEK 1981, S. 77ff; STEVEN 1998, S. 64, S. 68 und S. 79f; KLEINE 2002, S. 73f; CORSTEN 2007, S. 64; FANDEL 2007, S. 45ff.

ge verfügbar sind, z. B. aufgrund von Kapazitätsgrenzen der Betriebsmittel oder Beschaffungsengpässen bei Verbrauchsfaktoren. Zum anderen ist es möglich, dass von einem Produkt eine vorgegebene Mindestmenge hergestellt werden muss, z. B. aufgrund einer Lieferverpflichtung. Es sei b_m^I die Gütergrenze des m-ten Input mit $b_m^I \geqq 0$, $m = 1, ..., M$, und b_n^O die Mindestmenge des n-ten Output mit $b_n^O \geqq 0$, $n = 1, ..., N$, so lautet der Vektor für die Güterbeschränkung:

$$\mathbf{b} := \left(-b_1^I, ..., -b_M^I, b_1^O, ..., b_N^O\right)^T.$$

Führt man diese Güterbeschränkungen bei einer linearen Technologie ein, so ist

$$TM_L^\circ := \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \middle| \mathbf{z} = \sum_{j=1}^{J} \lambda_j \cdot \mathbf{z}_j, \mathbf{z} \geqq \mathbf{b}, \lambda_j \geqq 0, j = 1, .., J \right\}.$$

Die Technologie ist dann nur begrenzt größenproportional und begrenzt additiv.

2.3 Der Effizienzbegriff und ausgewählte Effizienzkonzepte

Die Technologiemenge beschreibt vollständig die Produktionsmöglichkeiten von Unternehmen im Sinne zulässiger Input-Output-Kombinationen.[35] Bei der Entscheidung, welche Produktionsalternative realisiert werden soll (ex ante-Betrachtung), aber auch bei der Beurteilung einer realisierten Produktion (ex post-Betrachtung), spielt das Effizienzkriterium eine wichtige Rolle. Eine Produktion $\mathbf{z} \in TM$ gilt als effizient, wenn es bei der gegebenen Technologiemenge TM keine Möglichkeit gibt, dieselbe bzw. eine höhere Ausbringungsmenge mit geringeren bzw. denselben Faktoreinsatzmengen zu erzeugen. Diese verbale Beschreibung der Effizienz kann durch folgende Definition formal präzise dargestellt werden:[36]

[35] Vgl. auch im Weiteren FANDEL 2007, S. 48ff.

[36] Zu den Definitionen der Effizienz und der Dominanz vgl. u. a. HILDENBRAND 1966, S. 66; STEUER 1986, S. 147ff; STEVEN 1998, S.10; SCHEEL 2000, S. 63; KLEINE 2002, S. 21; FANDEL 2007, S. 50.

Definition 2.1 ***(Effizienz):***
Eine Produktion $\mathbf{z} \in TM$ *heißt effizient, wenn keine andere Produktion* $\mathbf{z}' \in TM$ *existiert mit:*

$$\begin{array}{llll} & x'_m & \leqq & x_m \quad m = 1, ..., M, \\ & y'_n & \geqq & y_n \quad n = 1, ..., N \\ \textit{und} & x'_m & < & x_m \quad \textit{für mindestens ein } m \\ \textit{oder} & y'_n & > & y_n \quad \textit{für mindestens ein } n. \end{array}$$

In Vektorschreibweise: $\mathbf{z}' \geq \mathbf{z}$.

Diese Definition der Effizienz beinhaltet indirekt die Dominanzdefinition:

Definition 2.2 ***(Dominanz):***
Eine Produktion $\mathbf{z}' \in TM$ *dominiert eine Produktion* $\mathbf{z} \in TM$*, wenn gilt:* $\mathbf{z}' \geq \mathbf{z}$.

Die Dominanzdefinition stellt eine Umkehrung der Effizienzdefinition dar, sodass eine Produktion dann als effizient bezeichnet werden kann, wenn sie von keiner anderen Produktion aus der Technologiemenge dominiert wird.

Der hier eingeführte Effizienzbegriff wird auch als *Pareto-Koopmans-Effizienz* bezeichnet - kurz *PK-Effizienz*. Er basiert auf dem von PARETO 1906 formulierten wohlfahrtsökonomischen Prinzip, das seitdem die Basis für die Effizienzbegriffe bildet.[37] KOOPMANS 1951 erweiterte die wohlfahrtsökonomischen Überlegungen auf die Produktionsökonomie durch die produktionstheoretische Beschreibung effizienter Produktionen.[38] Da der Begriff der Pareto-Koopmans-Effizienz auf die Verwendung von Preisen für die Produktionsfaktoren und Produkte verzichtet, wird sie von FARRELL 1957 auch als *technische Effizienz* bezeichnet.[39]

Die Effizienz einer Produktion wird stets bezüglich der zugrunde liegenden Technologie ermittelt. Eine effiziente Produktion wird folglich durch keine andere Produktion aus der Technologiemenge dominiert. Ändert sich die Technologie, zum Beispiel aufgrund "technischen Fortschritts", so kann dies Auswirkungen auf die Effizienzbeurteilung einer Produktion haben.[40] Des Weiteren wird davon ausgegangen, dass die Outputs maximiert und die Inputs minimiert werden. Im Hinblick auf die Effizienz einer Produktion ist es folglich um so besser, je höher die Outputmengen und je geringer die Inputmengen sind.[41]

[37]Vgl. SCHEEL 2000, S. 62.
[38]Vgl. COOPER et al. 2006; S. xxii.
[39]Vgl. FARRELL 1957, S. 254.
[40]Vgl. SCHEEL 2000, S. 63.
[41]Vgl. DYCKHOFF 2006, S. 177.

Mit Hilfe des Effizienzkriteriums können eindeutig effiziente Produktionen von den ineffizienten unterschieden werden. Die Menge der effizienten Produktionen TM_{eff} ist diejenige Teilmenge der Technologiemenge ($TM_{\mathit{eff}} \subseteq TM$), die alle nicht dominierten Produktionen enthält. Formal lässt sie sich wie folgt darstellen:[42]

$$TM_{\mathit{eff}} := \left\{ \mathbf{z} \in TM \,\middle|\, \nexists \mathbf{z}' \in TM : \mathbf{z}' \geq \mathbf{z} \right\}.$$

Aus der Effizienzdefinition folgt, dass effiziente Produktionen nie im Inneren, sondern auf dem Rand einer Technologiemenge liegen, weshalb die Menge TM_{eff} auch als *effizienter Rand* bezeichnet wird.[43] Der Umkehrschluss, dass alle Produktionen auf dem Rand effizient sind, gilt jedoch nicht.

In der Literatur wurden auf Basis der Pareto-Koopmans-Effizienz weitere Effzienzkonzepte entwickelt. So wird die Effizienzdefinition in eine input- und eine outputorientierte Variante aufgespalten (*orientierte Effizienz*). Eine Produktion gilt als inputorientiert effizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, bei mindestens einem Input eine geringere Faktormenge einzusetzen und dabei dieselbe oder eine höhere Ausbringungsmenge zu erzeugen. Eine Produktion gilt als outputorientiert effizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, bei mindestens einem Output eine höhere Ausbringungsmenge zu erzeugen, wobei dieselbe oder eine geringere Faktormenge eingesetzt wird. Ist eine Produktion sowohl input- als auch outputorientiert effizient, so ist sie (unorientiert) effizient.[44]

Definition 2.3 ***(Orientierte Effizienz):***[45]

- *Eine Produktion* $\mathbf{z} \in TM$ *heißt* inputorientiert effizient*, wenn keine Produktion* $\mathbf{z}' \in TM$ *existiert mit:*

  $$\mathbf{x}' \leq \mathbf{x} \textit{ und } \mathbf{y}' \geqq \mathbf{y}.$$

- *Eine Produktion* $\mathbf{z} \in TM$ *heißt* outputorientiert effizient*, wenn keine Produktion* $\mathbf{z}' \in TM$ *existiert mit:*

  $$\mathbf{x}' \leqq \mathbf{x} \textit{ und } \mathbf{y}' \geq \mathbf{y}.$$

[42] Vgl. ZELEWSKI 1993, S. 24f; SCHEEL 2000, S. 63.
[43] Vgl. auch im Weiteren FANDEL 2007, S. 50f.
[44] Vgl. SCHEEL 2000, S. 68; KLEINE 2002, S. 95.
[45] Vgl. KLEINE 2002, S. 94.

Eine Abschwächung der Pareto-Koopmans-Effizienz stellt die *schwache Effizienz* dar.[46] Eine Produktion gilt als schwach effizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, bei *allen* Outputs eine höhere Ausbringungsmenge zu erzeugen, wobei bei *allen* Inputs eine geringere Einsatzmenge verwendet wird.

Definition 2.4 ***(Schwache Effizienz):***
Eine Produktion $\mathbf{z} \in TM$ *heißt schwach effizient bezüglich der Technologiemenge* TM*, wenn keine andere Produktion* $\mathbf{z}' \in TM$ *existiert mit:*

$$\mathbf{x}' < \mathbf{x} \text{ und } \mathbf{y}' > \mathbf{y}$$

$$\text{oder kurz:} \quad \mathbf{z}' > \mathbf{z}.$$

Auch beim Konzept der schwachen Effizienz ist eine Aufspaltung in eine input- und eine outputorientierte Variante möglich. Auf die formale Darstellung wird hier jedoch verzichtet.[47]

Ein weiteres Effizienzkonzept ist die *radiale Effizienz*.[48] Eine Produktion gilt als radial effizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, die Ausbringungsmengen aller Outputs proportional zu erhöhen und gleichzeitig alle Faktoreinsatzmengen um denselben Faktor zu reduzieren.[49]

Definition 2.5 ***(Radiale Effizienz):***
Eine Produktion $\mathbf{z} \in TM$ *heißt radial effizient bezüglich der Technologiemenge* TM*, wenn kein* $\tau > 0$ *existiert mit:*

$$\begin{pmatrix} -(1-\tau)\cdot\mathbf{x} \\ (1+\tau)\cdot\mathbf{y} \end{pmatrix} \in TM.$$

Die radiale Effizienz ist insbesondere in der orientierten Variante gebräuchlich (*orientierte radiale Effizienz*).[50] Eine Produktion gilt als inputorientiert radial ef-

[46] Zur schwachen Effizienz vgl. STEUER 1986, S. 221; SCHEEL 2000, S. 66.
[47] Zu einer Definition der schwachen orientierten Effizienz vgl. FÄRE et al. 1985, S. 28; SCHEEL 2000, S. 67.
[48] Dieses Effizienzkonzept liegt dem von FARRELL 1957 verwendeten Effizienzmaß zu Grunde, wobei dieser auf eine Arbeit von DEBREU 1951 zurückgreift. Vgl. LOVELL 1993, S. 10; DARAIO/SIMAR 2007, S. 24. Daher wird das auf der radialen Effizienz basierende Effizienzmaß auch als Debreu-Farrell Maß bezeichnet.
[49] Zur radialen Effizienz vgl. SCHEEL 2000, S. 66f.
[50] Zur orientierten radialen Effizienz vgl. FÄRE et al. 1985, S.53ff und S. 83ff; LOVELL 1993, S. 9ff; SCHEEL 2000, S. 66ff.

fizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, alle Inputs um denselben Faktor proportional zu verringern, und dabei dieselbe oder eine höhere Ausbringungsmenge zu erzeugen. Eine Produktion gilt als outputorientiert radial effizient, wenn es bei gegebener Technologiemenge keine Möglichkeit gibt, mit demselben oder einem geringeren Faktoreinsatz alle Outputs um denselben Faktor proportional zu erhöhen.[51]

Definition 2.6 ***(Orientierte radiale Effizienz):***

- *Eine Produktion* $\mathbf{z} \in TM$ *ist* inputorientiert radial effizient*, falls kein* $\theta < 1$ *existiert mit:*

$$\begin{pmatrix} -\theta \cdot \mathbf{x} \\ \mathbf{y} \end{pmatrix} \in TM.$$

- *Eine Produktion* $\mathbf{z} \in TM$ *ist* outputorientiert radial effizient*, falls kein* $\phi > 1$ *existiert mit:*

$$\begin{pmatrix} -\mathbf{x} \\ \phi \cdot \mathbf{y} \end{pmatrix} \in TM.$$

Mit den vorgestellten Effizienzdefinitionen können effiziente von ineffizienten Produktionen unterschieden werden. Soll die Effizienzbeurteilung darüber hinausgehen, so müssen graduelle Unterschiede zwischen den Produktionen gemessen werden,[52] sodass eine Rangbildung der Produktionen möglich ist. Dazu sind geeignete *Effizienzmaße* notwendig.

Die Aufgabe von Effizienzmaßen wird an einem Beispiel veranschaulicht. Gegeben ist eine Technologie mit zwei Inputs (mit den Inputquantitäten x_1 und x_2) und einem auf eine Mengeneinheit normierten Output ($\overline{y} = 1$). Die Technologiemenge ist als graue Fläche in Abbildung 2.6 dargestellt, ihr effizienter Rand ist durch die schwarze Linie gekennzeichnet. Ebenso in der Abbildung enthalten sind die vier Produktionen A, B, C und D. Einzig Produktion A befindet sich auf dem effizienten Rand der Technologiemenge und ist somit effizient. Dagegen liegen die Produktionen B, C und D im Inneren der Technologiemenge. In der Darstellung wird ersichtlich, dass Produktion C von Produktion B dominiert wird, da letztere bei gleicher Outputmenge geringere Inputmengen bei beiden Inputs einsetzt.

[51] Zum Unterschied von radialer und hyperbolischer Effizienz vgl. SCHEEL 2000, S. 67, Fußnote 56.
[52] Vgl. SCHEEL 2000, S. 75.

Produktion B ist also durch eine höhere Effizienz gekennzeichnet wie Produktion C. Ein ähnlicher Vergleich mit der Produktion D ist nicht möglich, da zwischen dieser und den anderen ineffizienten Produktionen keine Dominanzbeziehung besteht, bzw. da bei Produktion D von der Inputmenge x_1 zwar weniger, von der Inputmenge x_2 aber mehr eingesetzt wird als bei den Produktionen B und C.

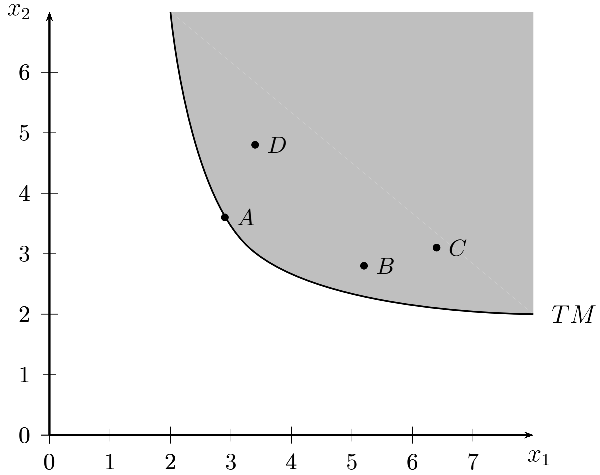

Abb. 2.6: Überlegungen zum Effizienzmaß

Mit Hilfe eines Effizienzmaßes sollen Aussagen über das unterschiedliche Ausmaß der Ineffizienz von Produktionen gemacht werden. Ein Effizienzmaß ist eine Funktion, die jeder Produktion unter Berücksichtigung der Technologiemenge einen reellwertigen Effizienzwert zuweist.[53] Dieser Effizienzwert soll die Diskrepanz zwischen der zu beurteilenden Produktion und einer effizienten Produktion (Benchmark) aus der Technologiemenge beschreiben.

In der Literatur werden verschiedene Eigenschaften genannt, die für ein ideales Effizienzmaß wünschenswert sind, u. a.:

- *Relevanz*: das Effizienzmaß soll alle relevanten Objekte (Inputs und Outputs) berücksichtigen,[54]

- *Skaleninvarianz*: das Effizienzmaß soll unabhängig von den Maßeinheiten der

[53] Vgl. STEINMANN 2002, S. 18.
[54] Vgl. DYCKHOFF 2006, S. 157.

Inputs und Outputs sein,[55]

- *Effizienzindikation*: das Effizienzmaß soll anzeigen, ob eine Produktion effizient ist (gemäß der zu Grunde liegenden Effizienzdefinition),[56]
- *Kompatibilität*: dominiert eine Produktion eine andere, muss sie einen höheren Effizienzwert aufweisen,[57]
- *einfache Berechenbarkeit*: der Rechenweg soll nachvollziehbar sein und[58]
- *einfache Interpretierbarkeit*: das Effizienzmaß soll Verbesserungspotenziale aufzeigen und quantifizieren.[59]

Ein Vergleich der Produktionen B und C aus Abbildung 2.6 zeigt, dass Produktionen, die nahe am effizienten Rand liegen, sich durch eine höhere Effizienz auszeichnen, als die weiter entfernt liegenden Produktionen. Daher ist es nahe liegend, den Abstand einer Produktion vom effizienten Rand als Maß für die Effizienz bzw. Ineffizienz heranzuziehen.[60] Allerdings gibt es unterschiedliche Möglichkeiten, einen Abstand zu messen, z. B. durch die Euklidische Metrik, die City Block-Metrik oder die Tschebyscheff-Metrik.[61]

2.4 Verfahren der Effizienzmessung

Die Effizienzmessung soll Aufschluss über die Effizienz einer realisierten Produktion unter der gegebenen Technologiemenge geben und somit die Leistung einer Entscheidungseinheit, z. B. eines Unternehmens, einer Abteilung, beurteilen. Ist die Produktion nicht effizient, so ist das Ausmaß der Ineffizienz von Interesse. Die Verfahren der Effizienzmessung müssen daher folgende Aufgaben lösen:

- Die Technologiemenge ist in der Praxis häufig nicht bekannt. Es liegen lediglich beobachtete Produktionen vor. Daher muss die Technologiemenge empirisch geschätzt werden.

[55] Vgl. STEINMANN 2002, S. 18f; DYCKHOFF 2006, S. 157.
[56] Vgl. STEINMANN 2002, S. 18f.
[57] Vgl. STEINMANN 2002, S. 18f; DYCKHOFF 2006, S. 157.
[58] Vgl. DYCKHOFF 2006, S. 157; BURGER 2008, S. 17.
[59] Vgl. BURGER 2008, S. 17.
[60] Vgl. KLEINE 2002, S. 179; DYCKHOFF 2006, S. 156
[61] Vgl. u. a. DYCKHOFF 2006, S. 156.

- Um zu bestimmen, ob eine Produktion effizient bezüglich der geschätzten Technologiemenge ist, und gegebenenfalls deren Grad der Ineffizienz zu ermitteln, wird ein geeignetes Effizienzmaß benötigt.

Seit Farrells Arbeit zur Messung der technischen Effizienz[62] wurden unterschiedliche Ansätze zur Bestimmung der Technologiemenge bzw. ihres effizienten Randes sowie zur Berechnung eines Effizienzwertes entwickelt. DARAIO/SIMAR 2007 klassifizieren die Verfahren der Effizienzmessung nach folgenden Kriterien:[63]

1) *Konstruktion der Technologiemenge:*

   Bei *parametrischen Verfahren* wird ein funktionaler Zusammenhang zwischen Input und Output angenommen.[64] Der Output y ist in der Regel univariat. Die Technologiemenge wird durch eine Randfunktion ($y = g(\mathbf{x}, \beta)$) definiert, die von den Inputs $\mathbf{x}$ und den Parametern β abhängt, z. B. in Form einer Cobb-Douglas-Funktion. Die Parameter werden auf Grundlage der beobachteten Daten mittels ökonometrischer Methoden geschätzt.[65] Die wesentlichen Vorteile parametrischer Verfahren sind die ökonomische Interpretierbarkeit der geschätzten Parameter sowie deren statistische Eigenschaften. Die Berücksichtigung von Datenschwankungen, z. B. aufgrund stochastischer Störeinflüsse oder Messfehler, ist zudem relativ einfach.[66] Schwierig ist die Wahl einer geeigneten Randfunktion, aber auch die gleichzeitige Berücksichtigung mehrerer Inputs und Outputs, da es hierfür multivariater parametrischer Verfahren bedarf.[67] Problematisch ist zudem, dass zur Schätzung der Parameter der Randfunktion nicht nur die effizienten, sondern auch die ineffizienten Produktionen herangezogen werden, was zu Verzerrungen der Ergebnisse führen kann.[68]

   Bei *nicht-parametrischen Verfahren* wird der Zusammenhang zwischen Inputs und Outputs als *Black Box* angesehen.[69] Der Rand der Technologiemenge wird nicht durch eine Funktion geschätzt, sondern auf Basis von beobachteten Produktionen mittels eines stückweise linearen Funktionsverlaufes approximiert.[70] Diese Approximation wird so gewählt, dass sie die beobachteten

[62] Vgl. FARRELL 1957.
[63] Vgl. DARAIO/SIMAR 2007, S. 26.
[64] Vgl. auch im Weiteren DARAIO/SIMAR 2007, S. 26.
[65] Zur Schätzung von Randfunktionen mittels ökonometrischer Methoden vgl. u. a. GREENE 1993; COELLI et al. 2005, S. 209ff; LAST/WETZEL 2009, S. 11ff.
[66] Vgl. SCHEEL 2000, S. 51; COELLI et al. 2005, S. 312.
[67] Vgl. COELLI et al. 2005, S. 312; DARAIO/SIMAR 2007, S. 26.
[68] Vgl. SCHEEL 2000, S. 52.
[69] Vgl. SCHEEL 2000, S. 50.
[70] Vgl. auch im Weiteren CANTNER et al. 2007, S. 22.

Produktionen möglichst eng umhüllt, wobei ausschließlich die Lage der Produktionspunkte entscheidend ist. Somit sind diese Verfahren weniger anfällig dafür, eine ungeeignete Form für die Randfunktion zu wählen.[71] Auch die gleichzeitige Berücksichtigung mehrerer Inputs und Outputs ist unproblematisch. Dagegen gestaltet sich die Berücksichtigung stochastischer Einflüsse als schwieriger.

2) *Berücksichtigung von Datenunsicherheit*

Bei *deterministischen Verfahren* wird angenommen, dass alle beobachteten Produktionen zur Technologiemenge gehören. Zufallseinflüsse sowie statistische Messfehler und Datenausreißer bleiben unberücksichtigt. Somit werden stochastische Einflüsse mit der Ineffizienz vermischt,[72] was zu verzerrten Ergebnissen bei der Effizienzmessung führen kann.

Bei *stochastischen Verfahren* werden Zufallseinflüsse in den Daten explizit berücksichtigt. Einzelne Beobachtungen können daher außerhalb der Technologiemenge liegen.[73] Die Schwierigkeit besteht darin, Zufallseinflüsse von der Ineffizienz zu trennen. Hierfür sind Verteilungsannahmen für einen Fehlerterm zur Abbildung der Zufallseinflüsse notwendig; bei einigen Modellen auch für den Ineffizienzterm.[74]

3) *Art der analysierten Daten*

Bei *statischen Verfahren* wird eine einperiodige Betrachtung vorgenommen, d. h. für jede der J zu untersuchenden Einheiten liegt nur je eine beobachtete Produktion vor, sodass folgende Daten der Effizienzmessung zugrunde gelegt werden:[75]

$$\mathscr{X} = \{\mathbf{z}_j | j = 1, ..., J\}.$$

Bei *dynamischen Verfahren* erfolgt eine mehrperiodige Betrachtung, d. h. für jede der J zu untersuchenden Einheiten liegen beobachtete Produktionen für T Perioden vor, sodass

$$\mathscr{X} = \{\mathbf{z}_{jt} | j = 1, ..., J, t = 1, ..., T\}.$$

[71] Vgl. LOVELL 1993, S. 19.
[72] Vgl. LOVELL 1993, S. 19.
[73] Vgl. auch im Weiteren DARAIO/SIMAR 2007, S. 28.
[74] Vgl. COELLI et al. 2005, S. 312; HERRERO 2005, S. 258.
[75] Vgl. auch im Weiteren DARAIO/SIMAR 2007, S. 28.

Somit ermöglichen diese Verfahren die Messung von Effizienzänderungen im Zeitablauf sowie die Schätzung des technischem Fortschritts. Da im Gegensatz zu statischen Verfahren für jede Einheit mehr als eine Beobachtung vorliegt, führen dynamische Verfahren zudem zu einer besseren Schätzung der Effizienz der einzelnen Einheiten.[76]

Die am weitesten verbreiteten Verfahren der Effizienzmessung sind der *Stochastic Frontier Approach (SFA)* sowie die *Data Envelopment Analysis (DEA)*.[77] Beim *SFA* wird eine stochastische Randfunktion geschätzt, sodass es sich um ein parametrisches, stochastisches Verfahren handelt.[78] Dagegen ist die *DEA* ein nichtparametrisches Verfahren.[79]

2.5 Data Envelopment Analysis

2.5.1 Hintergründe und Grundbegriffe

Die *Data Envelopment Analysis* - kurz *DEA* - ist ein nicht-parametrisches Verfahren der Effizienzmessung. Das ursprüngliche DEA-Modell ist deterministisch und statisch. Zahlreiche Erweiterungen existieren jedoch, die auch stochastische Einflüsse berücksichtigen oder mehrperiodige Betrachtungen ermöglichen.[80] Die DEA zeichnet sich durch ihre empirische Ausrichtung aus und bedarf nicht der zahlreichen a priori Annahmen, die bei anderen Ansätzen der Effizienzmessung notwendig sind.[81]

Die DEA geht auf eine Arbeit von FARRELL 1957 zurück, in der er das Ziel verfolgte, bessere Methoden zur Produktivitätsbeurteilung zu entwickeln.[82] Dazu setzte er einen aktivitätsanalytischen Ansatz ein. Farrells Effizienzkonzept ist eine

[76] Vgl. LOVELL 1993, S. 25.

[77] Vgl. HERRERO 2005, S. 257; DARAIO/SIMAR 2007, S. 29; LAST/WETZEL 2009, S. 14; KALB 2010, S. 52.

[78] Zu einer ausführlichen Darstellung des Stochastic Frontier Approach vgl. COELLI et al. 2005, S. 241ff.

[79] Weitere Verfahren der Effizienzmessung werden beschrieben in LOVELL 1993; HERRERO 2005; DARAIO/SIMAR 2007, S. 29; HACKMAN 2008; LAST/WETZEL 2009.

[80] Beispiele für stochastische DEA-Modelle sind LAND et al. 1993; SENGUPTA 1995, S. 133ff; SENGUPTA 1996; GONG/SUN 1998; POST 2001; POST et al. 2002; COOPER et al. 2011a; LEE/JOHNSON 2014. Beispiele für mehrperiodige DEA-Modelle sind SENGUPTA 1995, S. 133ff; LOVELL 1996, RAY 2004, S. 274ff; FÄRE et al. 2011; KAO/LIU 2014.

[81] Vgl. COOPER et al. 2011b, S. 2.

[82] Zur Historie der DEA vgl. u. a. ALI/SEIFORD 1993, S. 120; RAY 2004, S. 1ff; COOPER et al. 2011b, S. 3ff.

von Preisen unabhängige Betrachtung von Mengenrelationen (Input- und Outputmengen), bei der mehrere Inputs gleichzeitig berücksichtigt werden. Farrell beschrieb die Technologiemenge durch eine stückweise lineare, konvexe Hülle der Input-Output-Vektoren (Aktivitäten). Farrells Ansatz wurde von CHARNES et al. 1978 und 1981 für eine Technologie mit mehreren Outputs erweitert und als lineares Programm formuliert. Sie führten auch den Begriff *Data Envelopment Analysis* ein. Seither sind viele Arbeiten erschienen, in denen dieses DEA-Modell weiterentwickelt wurde; in zahlreichen Studien wurden DEA-Modelle verwendet.[83]

Zunächst wurde die DEA insbesondere im Non Profit-Bereich eingesetzt.[84] Das Verfahren wurde zunehmend vom privaten Sektor übernommen, sodass ein breites Spektrum an Anwendungsfeldern dazugekommen ist. Aktuelle Studien finden sich insbesondere im Dienstleistungssektor (z. B. in Banken, Krankenhäusern, Bildungseinrichtungen), für den die DEA als besonders geeignet gilt. Begründet wird dies damit, dass bei diesem Verfahren keine parametrische Produktionsfunktion vorgegeben werden muss.

Die Objekte, deren Effizienz bei der DEA untersucht wird, werden als Entscheidungseinheiten bzw. *Decision Making Units (DMUs)* bezeichnet. Es handelt sich um Einheiten mit Entscheidungskompetenz, das heißt, dass diese Einheiten auch verantwortlich für die zu bewertenden Produktionen sind.[85] Der Begriff Entscheidungseinheit bzw. DMU wird verwendet, um eine hohe Flexibilität bei den Anwendungsmöglichkeiten zu schaffen. Beispiele für DMUs sind Banken, Supermärkte, Schulen und Bibliotheken. DMUs können ganze Unternehmen oder Institutionen bezeichnen, sowie deren Teilbereiche, z. B. Werke und Abteilungen. Aber auch Regionen oder ganze Volkswirtschaften können als DMUs untersucht werden.

Die Beurteilung der Effizienz einer DMU erfolgt im Vergleich mit anderen DMUs. Eine Vergleichsgruppe umfasst J DMUs. Jede DMU_j, $j = 1, ..., J$, wird charakterisiert durch eine Produktion $\mathbf{z}_j$, die von ihr realisiert wurde. Die Produktionen aller DMUs der Vergleichsgruppe weisen dieselbe Struktur auf, d. h. jede der J Einheiten verwendet dieselben M Inputarten $(m = 1, ..., M)$, um daraus dieselben N Outputarten $(n = 1, ..., N)$ zu erzeugen.[86] Die Produktionen unterscheiden sich lediglich durch die Input- und Outputquantitäten. Kenntnisse der zugrunde liegenden Produktionsprozesse sind für die Effizienzmessung nicht not-

[83] Beispiele für DEA-Anwendungen finden sich in COOK/ZHU 2005, LAWRENCE/KLEINMAN 2009, COOPER et al. 2011c, BRUHN/HADWICH 2011b.

[84] Vgl. DYCKHOFF 2006, S. 176.

[85] Vgl. auch zum Folgesatz COOPER et al. 2006, S. 22.

[86] Vgl. COOPER et al. 2006, S. xix; DYCKHOFF 2006, S. 177.

wendig, vielmehr wird jede DMU als *Black Box* angesehen.[87]

Eine Produktion kann sowohl materielle als auch immaterielle Inputs und Outputs erfassen. Die Berücksichtigung eines Input oder Output ist allerdings nur dann möglich, wenn er in geeigneter Weise quantifizierbar und mindestens ordinal skalierbar ist.[88] Dies setzt voraus, dass eine geeignete Messvorschrift existiert, durch welche der Umfang des eingesetzten Input bzw. des produzierten Output in Form einer reellen Zahl abgebildet werden kann.[89] So wird der Bedarf an Werkstoffen durch physikalische Maße, zum Beispiel durch Stückzahlen, Gewichte oder Längen- und Flächenmaße erfasst; die Betriebsmittelnutzung kann in Zeiteinheiten erfasst werden.

2.5.2 Konstruktion der DEA-Technologiemenge

Um die Effizienz einer Entscheidungseinheit zu bestimmen, wird ihre realisierte Produktion mit den möglichen Produktionen aus der Technologiemenge verglichen. Die zu beurteilende Produktion wird dann als effizient bezeichnet, wenn sie durch keine Produktion aus der Technologiemenge dominiert wird. Voraussetzung für die Beurteilung einer Produktion ist also die Kenntnis der Technologiemenge. In der Praxis ist diese Technologiemenge meist nicht bekannt. Für die Effizienzmessung ist daher eine Approximation der wahren Technologiemenge notwendig. In der DEA sind die Grundlagen für diese Approximation

1) die beobachteten Produktionen aller Entscheidungseinheiten aus der Vergleichsgruppe sowie
2) die Annahmen zu den Eigenschaften der Technologiemenge.[90]

Die Konstruktion der DEA-Technologiemenge erfolgt nach dem *Prinzip der minimalen Extrapolation*, d. h. die Technologiemenge wird konstruiert als die kleinstmögliche Menge, die alle Annahmen erfüllt.[91] Wie gut die approximierte Technologiemenge die wahre Technologiemenge abbildet, hängt von der Menge und Qualität der verwendeten Daten sowie der Wahl der Annahmen ab.[92]

[87] Vgl. SCHEEL 2000, S. 2.
[88] Vgl. KLEINE 2002, S. 130.
[89] Vgl. auch im Weiteren STEVEN 1998, S. 7.
[90] Vgl. SCHEEL 2000, S. 40.
[91] Vgl. BANKER et al. 1984, S. 1081.
[92] Vgl. HACKMAN 2008, S. 2.

Die Menge der beobachteten Produktionen der J Entscheidungseinheiten aus der Vergleichsgruppe ist gegeben durch

$$\mathscr{X} := \{\mathbf{z}_j | j = 1, ..., J\} \quad \text{mit} \quad \mathscr{X} \in \mathbb{R}^{M+N}_{-+}. \tag{2.1}$$

Folgende Annahmen werden im Weiteren den DEA-Technologiemengen zugrunde gelegt:[93]

$i)$ *Empirische Vollständigkeit*
Alle beobachteten Produktionen sind zulässig, d. h. sie gehören zur Technologiemenge:

$$\mathbf{z}_j \in TM, \quad j = 1, ..., J.$$

$ii)$ *Konvexität*[94]
Jede Konvexkombination von Produktionen aus der Technologiemenge ist zulässig:[95]

$$\mathbf{z}_1, \mathbf{z}_2 \in TM, \quad \lambda_1 + \lambda_2 = 1, \quad \lambda_1, \lambda_2 \geqq 0 \quad \Rightarrow \quad \lambda_1 \mathbf{z}_1 + \lambda_2 \mathbf{z}_2 \in TM.$$

$iii)$ *Verschwendbarkeit (Ineffizienzpostulat)*[96]
Jede von einer Produktion aus der Technologiemenge dominierte Produktion ist zulässig:

$$\mathbf{z} \in TM, \quad \bar{\mathbf{z}} \leq \mathbf{z} \quad \Rightarrow \quad \bar{\mathbf{z}} \in TM,$$

wobei zwischen Verschwendbarkeit der Inputs und Verschwendbarkeit der Outputs unterschieden werden kann (s. Abschnitt 2.2.2, S. 16).

$iv)$ *Art der Skalenerträge*[97]
Die Art der Skalenerträge wird durch die Anforderungen an die Linearfaktoren λ_j, $j = 1, ..., J$, bestimmt. In der DEA werden insbesondere Modelle mit konstanten und variablen Skalenerträgen eingesetzt, auf welche sich diese Arbeit beschränkt:[98]

[93] Zu den Annahmen für die DEA-Technologiemenge vgl. u. a. BANKER et al. 1984, S. 1081; OLESEN/PETERSEN 1995, S. 447; RAY 2004, S. 27; COOPER et al. 2006, S. 42.
[94] Zur Konvexität vgl. Abschnitt 2.2.2, S. 16.
[95] Zu nicht konvexen Technologien, insbesondere der FDH-Technologie (*Free Disposal Hull*), vgl. u. a. KLEINE 2002, S. 129ff; DARAIO/SIMAR 2007, S. 33ff.
[96] Zur Verschwendbarkeit vgl. Abschnitt 2.2.2, S. 16.
[97] Zur Skalenvariation vgl. Abschnitt 2.2.2, S. 14.
[98] Zu Technologien mit nicht zunehmenden Skalenerträgen (*NIRS; non increasing returns to scale*)

- *konstante Skalenerträge:*

$$\begin{array}{ll} \mathbf{z}_j \in TM, j = 1, ..., J, & \\ \lambda_j \geqq 0, j = 1, ..., J & \end{array} \quad \Rightarrow \quad \sum_{j=1}^{J} \lambda_j \mathbf{z}_j \in TM,$$

- *variable Skalenerträge:*

$$\begin{array}{l} \mathbf{z}_j \in TM, j = 1, ..., J, \\ \sum\limits_{j=1}^{J} \lambda_j = 1, \\ \lambda_j \geqq 0, j = 1, ..., J \end{array} \quad \Rightarrow \quad \sum_{j=1}^{J} \lambda_j \mathbf{z}_j \in TM.$$

Als Zentrale Annahme für die DEA-Technologiemenge gilt die empirische Vollständigkeit, die SCHEEL 2000 auch als identitätsstiftend bezeichnet.[99] Bedeutend ist, dass diese Annahme die Zugehörigkeit von Beobachtungen zur Technologiemenge betrifft, wohingegen durch die Annahmen *ii)*, *iii)* und *iv)* festgelegt wird, welche virtuellen (nicht beobachteten) Produktionen aus der Menge der Beobachtungen abgeleitet werden.

Die Annahme der Verschwendbarkeit der Inputs und Outputs wird von BANKER et al. 1984 auch als Ineffizienzpostulat bezeichnet.[100] Dies soll zum Ausdruck bringen, dass eine ineffiziente Produktion - in Form von mehr Input oder weniger Output oder beidem - immer möglich und somit zulässig ist. Allerdings ist eine unbegrenzte Zulässigkeit der Verschwendbarkeit der Inputs bzw. der Outputs (im Sinne einer Faktorverschwendung bzw. Gütervernichtung) aus ökonomischer und ökologischer Sicht fragwürdig.[101] Da bei der Effizienzanalyse in der Regel nur effiziente Produktionen als Vergleichseinheiten herangezogen werden, hat die Zulässigkeit aller dominierten Produktionen aber auch keine Auswirkung auf die Effizienzbeurteilung einer DMU.

Wendet man das Prinzip der minimalen Extrapolation auf Basis der Annahmen der Vollständigkeit, Konvexität und Verschwendbarkeit an, so ergibt sich die DEA-Technologiemenge durch eine konvexe Umhüllung aller Produktionen $\mathbf{z}_j$,

und nicht abnehmenden Skalenerträgen (*NDRS; non decreasing returns to scale*) vgl. u. a. KLEINE 2002, S. 129ff; DARAIO/SIMAR 2007, S. 31ff.

[99] Vgl. SCHEEL 2000, S. 52.

[100] Vgl. BANKER et al. 1984, S. 1081.

[101] Vgl. auch im Weiteren KLEINE 2002, S. 143.

$j = 1,...,J$ (*Data Envelopment*). Die Technologiemenge ist somit

$$TM^{VRS} := \left\{ \mathbf{z} \in \mathbb{R}^{M+N}_{-+} \middle| \mathbf{z} \leqq \sum_{j=1}^{J} \lambda_j \mathbf{z}_j, \sum_{j=1}^{J} \lambda_j = 1, \lambda_j \geqq 0, j = 1,...,J \right\},$$

wobei die Bezeichnung TM^{VRS} darauf hinweist, dass es sich um eine Technologiemenge mit variablen Skalenerträgen (*variable returns to scale*) handelt. Die Technologiemenge TM^{VRS} ist die kleinstmögliche konvexe Menge, die alle beobachteten Produktionen enthält.[102] Eine Technologie mit konstanten Skalenerträgen (*constant returns to scale*) erhält man, wenn die Bedingung $\sum_{j=1}^{J} \lambda_j = 1$ aufgehoben wird, sodass auch alle nichtnegativen Vielfachen der beobachteten Produktionen zur Technologiemenge gehören:

$$TM^{CRS} := \left\{ \mathbf{z} \in \mathbb{R}^{M+N}_{-+} \middle| \mathbf{z} \leqq \sum_{j=1}^{J} \lambda_j \mathbf{z}_j, \quad \lambda_j \geqq 0, j = 1,...,J \right\}.$$

Diese Technologiemenge ist durch die Eigenschaften der Additivität sowie der Größenproportionalität gekennzeichnet. Sie entspricht somit der in Abschnitt 2.2.2 eingeführten linearen Technologie ($TM^{CRS} = TM_L^V$).[103]

Abbildung 2.7 zeigt jeweils eine DEA-Technologiemenge mit konstanten bzw. variablen Skalenerträgen, die auf Basis von fünf beobachteten Produktionen $A - E$ mit einem Input x und einem Output y konstruiert wurde.

Mit Ausnahme der Zulässigkeit des Produktionsstillstandes erfüllen die DEA-Technologiemengen die in Abschnitt 2.2.1 dargelegten Grundannahmen zu den Technologiemengen. Dies sind die Abgeschlossenheit, die Möglichkeit ertragreicher Produktion, die Unmöglichkeit des Schlaraffenlandes und die Unumkehrbarkeit der Produktion. Der Produktionsstillstand gehört nur unter spezifischen Annahmen bezüglich der Skalenerträge zur Technologiemenge, z. B. bei konstanten Skalenerträgen. Für die Effizienzbeurteilung spielt der Produktionsstillstand (Nullpunkt) jedoch keine Rolle, da er eine zulässige Produktion mit positiven Outputmengen nicht dominieren kann.[104]

Für die Effizienzmessung mit der DEA ist auch die zu den vorgestellten Technologiemengen duale Menge von Bedeutung.[105] Es sei P^{CRS} die zu TM^{CRS} duale Menge. Um die Dualität aufzuzeigen, werden im Folgenden die Technolo-

[102] Vgl. DARAIO/SIMAR 2007, S. 31.
[103] Vgl. DYCKHOFF/ALLEN 1997, S. 9; KLEINE 2002, S. 131.
[104] Vgl. ROSSMY 2007, S. 52.
[105] Zur Dualitätstheorie im Allgemeinen vgl. u. a. TURETSCHEK 1981, S. 15ff; STEUER 1986, S. 76ff; LUENBERGER/YE 2008, S. 79ff.

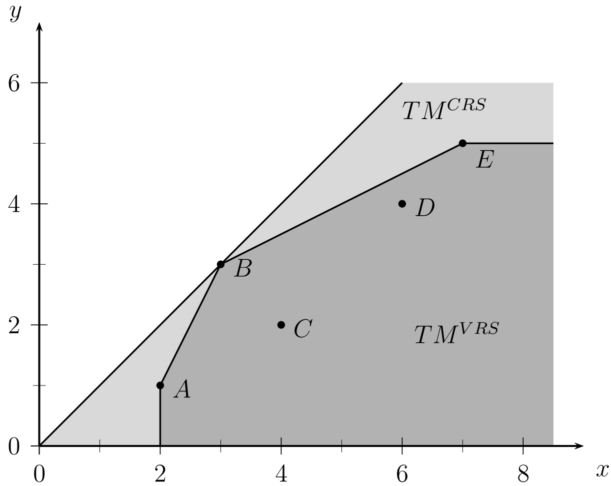

Abb. 2.7: DEA-Technologiemengen mit konstanten und variablen Skalenerträgen

giemenge TM^{CRS} und die Menge P^{CRS} simultan entwickelt, indem für beide die Annahmen *i)-iv)* zur DEA-Technologiemenge schrittweise eingeführt werden.[106] Des Weiteren sei TM_j der ausschließlich auf Basis der Produktion $\mathbf{z}_j$ konstruierte Teil der Technologiemenge und P_j die jeweils zu TM_j duale Menge. Als Dualvariablen dienen der Vektor der Input- und Output-Gewichte und die Variable w, wobei $\mathbf{u} := (u_1, .., u_M)^T$ der Vektor der Input-Gewichte und $\mathbf{v} := (v_1, ..., v_N)^T$ der Vektor der Output-Gewichte ist.

Annahme i) *empirische Vollständigkeit*

Die durch die j-te Beobachtung erzeugte Menge der Produktionsmöglichkeiten kann dargestellt werden als der Vektor der ersten $(M + N)$ Komponenten aus der Menge $TM_j(\lambda)$ für $\lambda = 1$, wobei

$$TM_j(\lambda) := \left\{ \left(\begin{array}{c} \mathbf{z} \\ \lambda \end{array} \right) \in \mathbb{R}_{-+}^{M+N+1} \middle| \mathbf{z} = \lambda_j \mathbf{z}_j, \lambda = \lambda_j, \lambda_j \geqq 0 \right\}, \ j = 1, ..., J.$$

Für $\lambda = 1$ enthält die Menge $TM_j(\lambda)$ ausschließlich die Beobachtung $\mathbf{z}_j$.

[106] Die folgende Herleitung der Menge P^{CRS} erfolgt in Anlehnung an OLESEN/ PETERSEN 1993, S. 13ff.

Es seien **u** die Dualvariablen der M Inputbedingungen in $TM_j(\lambda)$, **v** die Dualvariablen der N Outputbedingungen und w die Dualvariable für die Bedingung $\lambda = \lambda_j$. Dann ist die zu $TM_j(\lambda)$ duale Menge:

$$P_j := \left\{ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ w \end{pmatrix} \in \mathbb{R}^{M+N+1} \middle| \mathbf{v}^T \mathbf{y}_j - \mathbf{u}^T \mathbf{x}_j + w\lambda \leqq 0 \right\}, \quad j = 1, ..., J.$$

Die beiden Mengen $TM_j(\lambda)$ und P_j sind konvexe Kegel in $\mathbb{R}^{M+N+1}$. $TM_j(\lambda)$ ist der zu P_j polare Kegel:[107]

$$\begin{aligned} TM_j(\lambda) &= P_j{}^\circ \\ &= \left\{ \begin{pmatrix} \mathbf{z} \\ \lambda \end{pmatrix} \in \mathbb{R}_{-+}^{M+N+1} \middle| \forall \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ w \end{pmatrix} \in P_j : \mathbf{v}^T \mathbf{y} - \mathbf{u}^T \mathbf{x} + w\lambda \leqq 0 \right\}, \\ &\qquad j = 1, ..., J. \end{aligned} \tag{2.2}$$

Die Menge P_j ist ein Halbraum in $\mathbb{R}^{M+N+1}$ mit dem Normalenvektor $\mathbf{z}_j$. Alle Gewichte $(\mathbf{u}^T, \mathbf{v}^T, w)^T$ aus P_j erzeugen Halbräume in $\mathbb{R}^{M+N}$, welche die Produktion $\mathbf{z}_j$ enthalten. Die Menge $TM_j(\lambda = 1)$ ist die Schnittmenge aller Halbräume, welche die Produktion $\mathbf{z}_j$ enthalten und kann somit als eine Umhüllung dieser Produktion interpretiert werden.[108]

***Annahmen iii)** Verschwendbarkeit und **iv)** konstante Skalenerträge*

Die Einführung der Verschwendbarkeit und konstanter Skalenerträge führt zu einer Erweiterung der Produktionsmöglichkeiten, die durch die j-te Beobachtung erzeugt werden, sodass

$$TM_j^{CRS} := \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \middle| \mathbf{z} \leqq \lambda_j \mathbf{z}_j, \lambda_j \geqq 0 \right\}, \quad j = 1, ..., J.$$

[107] Zur Polarität vgl. HILDENBRAND 1966, S.73; GRITZMANN 2013, S. 292ff.

[108] Ein Halbraum ist eine durch eine Hyperebene begrenzte Menge eines Raumes. Zu den Definitionen eines Halbraumes und einer Hyperebene vgl. STEUER 1986, S. 36; LUENBERGER/YE 2008, S. 517ff; GRITZMANN 2013, S. 253ff.

Es sei P_j^{CRS} die zu TM_j^{CRS} duale Menge mit

$$P_j^{CRS} := \left\{ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \in \mathbb{R}_+^{M+N} \middle| \mathbf{v}^T \mathbf{y}_j - \mathbf{u}^T \mathbf{x}_j \leqq 0 \right\}, \quad j = 1, ..., J.$$

Aufgrund der Einführung konstanter Skalenerträge entfällt die Dualvariable w. Die Annahme der Verschwendbarkeit führt dazu, dass die Dualvariablen **u** und **v** nichtnegativ sind ($\mathbf{u} \in \mathbb{R}_+^M, \mathbf{v} \in \mathbb{R}_+^N$). Die Menge P_j^{CRS} ist ein Halbraum im nichtnegativen Orthant $\mathbb{R}_+^{M+N}$ mit dem Normalenvektor $\mathbf{z}_j$.

Annahme ii) *Konvexität*

Die Annahme der Konvexität erweitert die Produktionsmöglichkeiten um die Konvexkombinationen von Produktionen aus den Mengen TM_j^{CRS}, $j = 1, ..., J$. Die daraus resultierende Technologiemenge TM^{CRS} und der duale Kegel der Input-Output-Gewichte P^{CRS} können auf Basis der Mengen TM_j^{CRS} bzw. P_j^{CRS} erzeugt werden:

$$\begin{aligned} TM^{CRS} := \quad & conv\left(\bigcup_{j=1}^{J} TM_j^{CRS} \right) \\ = \quad & conv\left(\left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \middle| \mathbf{z} \leqq \lambda_j \mathbf{z}_j, \lambda_j \geqq 0 \right\}, j = 1, ..., J \right) \\ = \quad & \left\{ \mathbf{z} \in \mathbb{R}_{-+}^{M+N} \middle| \mathbf{z} \leqq \lambda_j \mathbf{z}_j, \lambda_j \geqq 0, j = 1, ..., J \right\}. \end{aligned} \tag{2.3}$$

Die Menge TM^{CRS} ist eine konvexe Umhüllung der J Technologiemengen TM_j, die durch die einzelnen Beobachtungen $\mathbf{z}_j$, $j = 1, ..., J$, erzeugt wurden.

$$\begin{aligned} P^{CRS} := \quad & \bigcap_{j=1}^{J} P_j^{CRS} \\ = \quad & \left\{ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \in \mathbb{R}_+^{M+N} \middle| \mathbf{v}^T \mathbf{y}_j - \mathbf{u}^T \mathbf{x}_j \leqq 0, j = 1, ..., J \right\}. \end{aligned} \tag{2.4}$$

Die Menge P^{CRS} ist die Schnittmenge aller Halbräume P_j^{CRS}, $j = 1, ..., J$, und somit ein lineares Ungleichungssystem.

Der Zusammenhang zwischen der Technologiemenge TM^{CRS} und der Menge P^{CRS} ist im folgenden Satz beschrieben:

Satz 2.1 [109] *Es sei der konvexe Kegel der Input-Output-Gewichte P^{CRS} gegeben durch (2.4). Die korrespondierende Technologiemenge TM^{CRS} in (2.3) ist die Schnittmenge des polaren Kegels $\left(P^{CRS}\right)^\circ$ mit dem Orthant $\mathbb{R}^{M+N}_{-+}$:*

$$\begin{aligned} TM^{CRS} = \quad & \left(P^{CRS}\right)^\circ \cap \mathbb{R}^{M+N}_{-+} \\ & = \left\{ \mathbf{z} \in \mathbb{R}^{M+N}_{-+} \middle| \forall \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \in P^{CRS} : \mathbf{v}^T\mathbf{y} - \mathbf{u}^T\mathbf{x} \leqq 0 \right\}. \end{aligned}$$

Entsprechend kann die zu TM^{VRS} duale Menge P^{VRS} hergeleitet werden. Diese unterscheidet sich von P^{CRS} dadurch, dass die Annahme konstanter Skalenerträge nicht gilt, sodass

$$P^{VRS} := \left\{ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ w \end{pmatrix} \in \mathbb{R}^{M+N}_{+} \times \mathbb{R} \middle| \mathbf{v}^T\mathbf{y}_j - \mathbf{u}^T\mathbf{x}_j + w \leqq 0, j = 1, ..., J \right\}.$$

2.5.3 *Entwicklung der DEA-Effizienzmaße und eines allgemeinen DEA-Modells*

Auf Basis der DEA-Technologiemenge wird die Effizienz der beobachteten Produktionen ermittelt. Mit Hilfe des Effizienzkriteriums[110] lassen sich bereits zwei Teilmengen von Produktionen unterscheiden: die Menge der effizienten Produktionen TM_{eff}, sowie die Menge der ineffizienten bzw. dominierten Produktionen $TM \setminus TM_{eff}$.[111] Da effiziente Produktionen auf dem Rand der Technologiemenge liegen, sind Aussagen über die Effizienz einer Produktion immer auch von der zugrunde gelegten Technologiemenge abhängig.[112]

Bei den ineffizienten Produktionen ist eine Beurteilung des Ausmaßes ihrer Ineffizienz von Interesse. Diese Beurteilung ermöglicht das DEA-Effizienzmaß. Dazu ermittelt dieses die Diskrepanz zwischen der zu beurteilenden Produktion und einer effizienten Produktion aus der Technologiemenge,[113] welche die zu beurteilende Produktion dominiert und als *effiziente Referenz* bezeichnet wird. In der DEA wird das Ausmaß der Diskrepanz zwischen diesen beiden Produktionen in

[109] Zum Satz und Beweis vgl. OLESEN/PETERSEN 1993, S. 15 (Theorem 2, leicht modifiziert) u. S. 31.
[110] Vgl. Abschnitt 2.3.
[111] Vgl. KLEINE 2002, S. 161.
[112] Vgl. KLEINE 2002, S. 160.
[113] Vgl. ALI/SEIFORD 1993, S. 135.

der Regel durch ein abstandsbasiertes Effizienzmaß beschrieben.[114] Je größer der Abstand, desto weniger effizient ist die zu beurteilende Produktion. Die DEA-Effizienzmaße unterscheiden sich darin, welche Produktion aus der Technologiemenge als effiziente Referenz gewählt wird und wie der Abstand zu dieser gemessen wird.

Es sei DMU_o die zu beurteilende Entscheidungseinheit mit $o \in \{1,...,J\}$, charakterisiert durch ihre realisierte Produktion $\mathbf{z}_o$, und $\hat{\mathbf{z}}$ deren effiziente Referenz. Die effiziente Referenz kann eine tatsächlich beobachtete Produktion oder eine virtuelle Produktion aus der Technologiemenge sein. Sie wird durch Linearkombination der beobachteten Produktionen aus der Vergleichsgruppe generiert mit

$$\hat{\mathbf{z}} = \begin{pmatrix} -\hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{pmatrix} := \begin{pmatrix} -\sum_{j=1}^{J} \lambda_j \cdot \mathbf{x}_j \\ \sum_{j=1}^{J} \lambda_j \cdot \mathbf{y}_j \end{pmatrix} = \sum_{j=1}^{J} \lambda_j \cdot \mathbf{z}_j .$$

Die Möglichkeiten bei der Bildung der Referenz werden durch die Annahmen bezüglich der Technologiemenge bestimmt. Die Linearfaktoren $\lambda_j, j = 1,..,J$, geben an, welchen Anteil die jeweilige Produktion $\mathbf{z}_j$ bei der Bildung der Referenz $\hat{\mathbf{z}}$ hat.

Der Abstand zwischen der Produktion $\mathbf{z}_o$ und deren Referenz $\hat{\mathbf{z}}$ berechnet sich aus den Abweichungen (Differenzen) bei den einzelnen Inputs und Outputs. Da die Outputmengen zu maximieren, die Inputmengen zu minimieren sind, ist bei den Abweichungen auch eine unterschiedliche Handhabung notwendig, weshalb unterschiedliche Abweichungsvariablen verwendet werden. Eine Abweichung bei einem Input liegt dann vor, wenn DMU_o eine größere Inputmenge einsetzt wie die Referenz (Überschreitung), wohingegen eine Abweichung beim Output besteht, wenn DMU_o eine geringere Ausbringungsmenge im Vergleich zur Referenz erzielt hat (Unterschreitung). Die Abweichungen sind immer nichtnegativ, da sonst die Referenz die Produktion $\mathbf{z}_o$ nicht dominieren würde. Es seien s_m^- die Überschreitung beim m-ten Input ($m = 1,...,M$) und s_n^+ die Unterschreitung beim n-ten Output ($n = 1,...,N$) mit

$$\begin{aligned}
&s_m^- := x_{om} - \hat{x}_m, && m = 1,...,M, \\
&s_n^+ := \hat{y}_n - y_{on}, && n = 1,...,N, \\
&s_m^-, s_n^+ \geqq 0, && m = 1,...,M,\ n = 1,...,N.
\end{aligned}$$

Die Abweichungen können zudem mit nichtnegativen Gewichten bewertet werden.

[114]Vgl. KLEINE 2002, S. 179.

Es sei t_m^- die Gewichtung der Unterschreitung beim m-ten Input ($m = 1,...,M$), mit $t_m^- \geqq 0$, und t_n^+ die Gewichtung der Überschreitung beim n-ten Output ($n = 1,...,N$), mit $t_n^+ \geqq 0$.

Es sei Γ das Maß für den Abstand zwischen der zu beurteilenden Produktion und der effizienten Referenz. Das Abstandsmaß Γ aggregiert die $M+N$ gewichteten Abweichungen zu einem reellen Wert ($\Gamma : \mathbb{R}_+^{M+N} \to \mathbb{R}$):[115]

$$\Gamma(t_1^- \cdot s_1^-,...,t_M^- \cdot s_M^-, t_1^+ \cdot s_1^+,...,t_N^+ \cdot s_N^+).$$

Die effiziente Referenz $\hat{\mathbf{z}}$ wird so bestimmt, dass der Abstand zur Produktion $\mathbf{z}_o$ maximiert wird. Somit lässt sich für die Berechnung der Effizienz einer DMU_o ein allgemeines DEA-Modell (DEA) formulieren:

$$\begin{array}{rll}
(DEA) & & \\
\max & \Gamma(t_1^- \cdot s_1^-,...,t_M^- \cdot s_M^-, t_1^+ \cdot s_1^+,...,t_N^+ \cdot s_N^+) & \\
\text{u.d.N.} & & \\
(1) & \hat{x}_m + s_m^- = x_{om}, & m = 1,...,M \\
(2) & \hat{y}_n - s_n^+ = y_{on}, & n = 1,...,N \\
(3) & s_m^-, s_n^+ \geqq 0, & m = 1,...,M,\ n = 1,...,N \\
(4) & (-\hat{x}_1,...,-\hat{x}_M, \hat{y}_1,...,\hat{y}_N)^T \in TM. &
\end{array}$$

Zur Ermittlung der Effizienz aller DMUs einer Vergleichsgruppe ist das Modell für jede der J DMUs zu lösen.

Aufgrund der Nichtnegativität der Abweichungsvariablen (Nebenbedingungen (3)) wird durch die Nebenbedingungen (1) und (2) gewährleistet, dass die Referenz die zu beurteilende Produktion dominiert. Das heißt, die Referenz ist in jedem Input und Output genauso gut oder besser, im Sinne von mehr Output oder weniger Input, als die zu beurteilende Produktion. Gibt es keine solche Referenz, sind die Abweichungsvariablen gleich null. Die zu beurteildende Produktion liegt auf dem Rand der Technologiemenge und ist effizient. Durch die Maximierung des Abstandsmaßes Γ wird nicht nur eine dominierende Referenz gewählt, sondern eine mit einem möglichst großen Abstand von der Produktion $\mathbf{z}_o$. Der Abstand ist, unabhängig vom gewählten Abstandsmaß, genau dann möglichst groß, wenn die Referenz auf dem Rand der Technologiemenge liegt.

Die zu beurteilende Produktion ist effizient bezüglich der Technologiemenge, wenn im Optimum sämtliche Werte der Abweichungsvariablen gleich null sind ($s_m^{-*} = 0,\ m = 1,...,M,\ s_n^{+*} = 0,\ n = 1,...,N$). In diesem Fall ist $\lambda_o = 1$ und

[115] Zur folgenden Herleitung der DEA-Zielfunktion vgl. KLEINE 2002, S. 179ff.

$\lambda_j = 0, \forall j \neq o$ und somit $\mathbf{z}_o = \hat{\mathbf{z}}$. Die Produktion $\mathbf{z}_o$ liegt dann auf dem effizienten Rand der Technologiemenge, sodass keine Diskrepanz zwischen dieser und der effizienten Referenz besteht. Voraussetzung für diese Beurteilung ist eine streng monoton steigende DEA-Zielfunktion. Im Falle einer monoton steigenden Zielfunktion kann es sich bei einem Zielfunktionswert von null sowohl um eine effiziente als auch um eine schwach effiziente Produktion handeln.[116]

Die zu beurteilende Produktion ist ineffizient bezüglich der Technologiemenge, wenn im Optimum mindestens eine Abweichungsvariable größer als null ist. In diesem Fall ist $\lambda_o = 0$ und es existiert eine effiziente Referenz, welche die Produktion $\mathbf{z}_o$ dominiert, sodass $\mathbf{z}_o \leq \hat{\mathbf{z}}$. Die Produktion $\mathbf{z}_o$ liegt dann im Inneren oder auf dem nicht PK-effizienten Rand der Technologiemenge. Zur Bestimmung eines Effizienzwertes für diese Produktion ist die DEA-Zielfunktion zu spezifizieren. Die DEA-Effizienzmaße unterscheiden sich in der Art der Aggregation der Abweichungen (Abstandsmaße) und in der Wahl der Gewichtungen.

Es gibt viele Möglichkeiten, den Abstand zwischen zwei Produktionspunkten zu messen. Die im Folgenden verwendeten Abstandsmaße beruhen auf den L_r-*Metriken*. Der L_r-Abstand zwischen einer zu beurteilenden Produktion und deren Referenz lautet:[117]

$$L_r = \left(\sum_{m=1}^{M} |x_{om} - \hat{x}_m|^r + \sum_{n=1}^{N} |y_{on} - \hat{y}_n|^r \right)^{\frac{1}{r}} = \left(\sum_{m=1}^{M} |s_m^-|^r + \sum_{n=1}^{N} |s_n^+|^r \right)^{\frac{1}{r}}$$

mit $r \in \{1,2,...\} \cup \{\infty\}$.

Für die DEA sind die additive Metrik (L_1-Metrik)[118] und die Maximin-Metrik (L_∞-Metrik)[119], sowie eine korrigierte Maximin-Metrik (ε-Maximin) von Bedeutung. In Abbildung 2.8 sind die entsprechenden DEA-Zielfunktionen enthalten.

Ein Effizienzmaß auf Basis der additiven Metrik bildet die Summe aller gewichteten Abweichungen von der effizienten Referenz. Dabei handelt es sich, für Gewichtungen $t_m^-, t_n^+ > 0, m = 1,...,M, n = 1,...,N$, um eine streng monoton steigende Zielfunktion, sodass bei einem optimalen Zielfunktionswert von null ($\Gamma^* = 0$) die zu beurteilende Produktion effizient bezüglich der Technologiemenge ist. Bei einer Maximin-Metrik ist dagegen ein optimaler Zielfunktionswert von null zwar ein Hinweis auf eine effiziente Produktion. Diese kann aber auch schwach effizient sein, da es sich um eine monoton, nicht aber streng monoton steigende Zielfunktion handelt. Durch die Addition der mit einer hinreichend klei-

[116] Zum Beweis siehe KLEINE 2002, S. 180 ff.
[117] Zu den L_r-Metriken vgl. u. a. STEUER 1986, S. 44ff.
[118] Auch: City-Block-Metrik.
[119] Auch: Tschebycheff-Metrik.

Metrik	DEA-Zielfunktion Γ
Additiv	$\sum_{m=1}^{M} t_m^- \cdot s_m^- + \sum_{n=1}^{N} t_n^+ \cdot s_n^+$
Maximin	$\min_{m,n}\{t_m^- \cdot s_m^-, t_n^+ \cdot s_n^+\}$
ε-Maximin	$\min_{m,n}\{t_m^- \cdot s_m^-, t_n^+ \cdot s_n^+\} + \varepsilon \cdot \left(\sum_{m=1}^{M} s_m^- + \sum_{n=1}^{N} s_n^+\right)$

Abb. 2.8: Ausgewählte Abstandsmaße als DEA-Zielfunktion

nen Zahl ε gewichteten Summe aller Abweichungen erhält man eine streng monoton steigende Zielfunktion, sodass bei einem optimalen Zielfunktionswert von null Effizienz vorliegt.

Die vorgestellten DEA-Effizienzmaße ermöglichen sowohl input- als auch outputorientierte Effizienzbetrachtungen. Während nichtorientierte Effizienzmaße die Abweichungen der Inputs und Outputs simultan erfassen, werden bei den orientierten Effizienzmaßen nur die Abweichungen der Inputs oder der Outputs erfasst, indem die nicht zu berücksichtigenden Kriterien auf null normiert werden ($t_m^- = 0$, $m = 1, ..., N$ bzw. $t_n^+ = 0$, $n = 1, ..., N$). Auf diese Weise ergeben sich die in Abbildung 2.9 dargestellten orientierten Zielfunktionen.

Metrik	Γ inputorientiert ($t_n^+ = 0$, $n = 1, ..., N$)	Γ outputorientiert ($t_m^- = 0$, $m = 1, ..., M$)
Additiv	$\sum_{m=1}^{M} t_m^- \cdot s_m^-$	$\sum_{n=1}^{N} t_n^+ \cdot s_n^+$
Maximin	$\min_{m}\{t_m^- \cdot s_m^-\}$	$\min_{n}\{t_n^+ \cdot s_n^+\}$
ε-Maximin	$\min_{m}\{t_m^- \cdot s_m^-\} + \varepsilon \cdot \sum_{m=1}^{M} s_m^-$	$\min_{n}\{t_n^+ \cdot s_n^+\} + \varepsilon \cdot \sum_{n=1}^{N} s_n^+$

Abb. 2.9: DEA-Zielfunktion - input- und outputorientiert

Im Falle einer Inputorientierung stellen aufgrund der Nebenbedingungen (1) aus

(DEA), S. 39, die Outputmengen der Produktion $\mathbf{z}_o$ untere Schranken für die Outputs der Referenz $\hat{\mathbf{z}}$ dar, im Falle einer Outputorientierung sind die Inputmengen von Produktion $\mathbf{z}_o$ obere Schranken für die Inputs der Referenz $\hat{\mathbf{z}}$ (aufgrund der Nebenbedingungen (2)). Eine positive Abweichung von der Schranke hat keine Auswirkungen auf den Zielfunktionswert.[120]

Die Gewichtungen ermöglichen nicht nur orientierte Effizienzbetrachtungen, sondern dienen vor allem der Normierung der Abweichungen. Bei *einfachen Gewichten* wird ein konstanter Wert als Normierungsfaktor für die Abweichungen gewählt. Allerdings ist ein solches Effizienzmaß nicht skaleninvariant, d. h. der errechnete Effizienzwert einer Produktion hängt davon ab, in welchen Einheiten, z. B. Kilogramm, Stunden, die Inputs und Outputs gemessen wurden.[121] Werden die Abweichungen ins Verhältnis zu geeigneten Bezugsgrößen gesetzt, können skaleninvariante Gewichte generiert werden, wie die Bandbreiten-Gewichte und die DMU-spezifischen Gewichte.[122]

Als Bandbreite beim m-ten Input bzw. n-ten Output (Δx_m bzw. Δy_n) wird die Differenz zwischen maximaler und minimaler Input- bzw. Outputquantität einer Technologiemenge bezeichnet. Diese Differenz ist von der zu Grunde liegenden Technologie abhängig. Bei den *Bandbreiten-Gewichten* erfolgt die Normierung, indem die Abweichung beim m-ten Input bzw. n-ten Output ins Verhältnis zur entsprechenden Bandbreite gesetzt wird.[123] Durch diese Art der Normierung wird aus der absoluten Abweichung eine prozentuale Abweichung in Bezug auf die Bandbreite, also die maximale Differenz des jeweiligen Input bzw. Output bezogen auf die Technologiemenge ($t_m^- := \frac{1}{\Delta x_m}$, $t_n^+ := \frac{1}{\Delta y_n}$).

Während bei Bandbreiten-Gewichten die Normierungen für alle Produktionen identisch sind, unterscheiden sich die Normierungen bei den *DMU-spezifischen Gewichten*. Bei diesen wird die Abweichung beim m-ten Input bzw. n-ten Output ins Verhältnis zur entsprechenden Input- bzw. Outputquantität der DMU_o gesetzt. Durch diese Art der Normierung wird die Abweichung als prozentuale Abweichung der Input- bzw. Outputquantitäten der zu beurteilenden Produktion von der Referenz erfasst ($t_m^- := \frac{1}{x_{om}}$, $t_n^+ := \frac{1}{y_{on}}$).

[120] Die Verwendung eines orientierten Effizienzmaßes bei einer additiven Metrik in der DEA-Zielfunktion führt allerdings dazu, dass diese, ebenso wie bei der Maximin-Metrik, nur monoton steigend ist. Auch hier kann durch die Einführung eines Korrekturterms in der Zielfunktion, d. h. die Addition der mit einer hinreichend kleinen Zahl ε gewichteten Summe aller Abweichungen, eine streng monoton steigende Zielfunktion erzeugt werden. Vgl. ROSSMY 2007, S. 56f.

[121] Vgl. u. a. KLEINE 2002, S. 185; LOVELL/PASTOR 1995.

[122] Neben den hier vorgestellten Gewichten sind auch andere Varianten der Normierung möglich. Vgl. KLEINE 2002, S. 190.

[123] Zu DEA-Modellen mit Bandbreitengewichten (auch: Range-Adjusted-Faktoren) vgl. COOPER et al. 1999, S. 18ff; KLEINE/SEBASTIAN 2005, S. 388.

Da Bandbreiten-Gewichte für alle DMUs gleich sind, hat dieselbe absolute Abweichung bei einem Input bzw. Output zweier verschiedener Produktionen dieselbe Auswirkung auf deren Effizienzwert. Bei DMU-spezifischen Gewichten kann dieselbe absolute Abweichung dagegen unterschiedliche Auswirkungen haben, je nach der von der jeweiligen DMU realisierten Input- bzw. Outputquantität.[124]

2.5.4 Ausgewählte DEA-Modelle: CCR-Modell und BCC-Modell

Das vorgestellte allgemeine DEA-Modell (DEA), S. 39, kann nun im Hinblick auf die Technologiemenge sowie das Effizienzmaß konkretisiert werden. So lässt sich eine Vielzahl an Modellen erzeugen, die auf verschiedenen Annahmen beruhen, daher aber auch zu unterschiedlichen Effizienzbeurteilungen führen können. Im Folgenden werden das CCR-Modell und das BCC-Modell hergeleitet. Diese beiden originären DEA-Modelle stellen die Basis für zahlreiche Weiterentwicklungen und Ergänzungen der DEA dar. Im Weiteren werden sie daher auch als DEA-Basismodelle bezeichnet.[125]

Beide Modelle verwenden ein inputorientiertes, radiales Effizienzmaß, das sich anhand des Maximin-Ansatzes mit DMU-spezifischen Gewichten herleiten lässt.[126] Die zu maximierende DEA-Zielfunktion $\Gamma(t_1^- \cdot s_1^-, ..., t_M^- \cdot s_M^-, t_1^+ \cdot s_1^+, ..., t_N^+ \cdot s_N^+)$ konkretisiert sich dann zu

$$\zeta := \min\left\{\frac{1}{x_{o1}} \cdot s_1^-, ..., \frac{1}{x_{oM}} \cdot s_M^-\right\}.$$

Das Maß ζ minimiert die prozentuale Abweichung der zu beurteilenden Produktion von deren Referenz über alle Inputs. Ausschlaggebend für den Wert von ζ ist folglich derjenige Input, bei welchem diese prozentuale Abweichung am geringsten ist. Dieses Maß gibt somit an, um wie viel Prozent der Inputvektor $\mathbf{x}_o$

[124] Weichen bei einer in Kilogramm gemessenen Ausbringungsmenge zwei verschiedene Produktionen jeweils um 1 Kilogramm von ihrer Referenz ab, so hat dies bei Bandbreitengewichte dieselbe Auswirkung auf den Zielfunktionswert (z. B. bei einer Bandbreite von 100 kg: 1/100). Bei den DMU-spezifischen Gewichten spielt jedoch die Outputquantität der jeweiligen DMU eine Rolle, d. h. tritt diese Abweichung bei der ersten Produktion bei einer Outputquantiät von 50 kg, bei der zweiten Produktion bei einer Outputquantität von 100 kg auf, so ergeben sich die gewichteten Abweichungen als 1/50 bzw. 1/100. Dies entspricht der prozentualen Abweichung von 2% bzw. 1%.

[125] Für weitere DEA-Modelle vgl. u. a. KLEINE 2002, S. 179ff; BANKER et al. 2011, S. 57ff; COOPER et al. 2006, S. 83ff.

[126] Zur folgenden Herleitung des inputorientierten, radialen Effizienzmaßes vgl. ALI/SEIFORD 1993, S. 134ff; KLEINE 2002, S. 200ff; ROSSMY 2007, S. 58ff. Die Herleitung eines outputorientierten, radialen Effizienzmaßes ist analog hierzu möglich.

reduziert werden muss, um die Produktion $\mathbf{z}_o$ schwach effizient zu machen bzw. die schwach effiziente Referenz zu erreichen. Durch die Maximierung von ζ wird die Referenz

$$\hat{\mathbf{z}} = \begin{pmatrix} -(1-\zeta)\cdot\mathbf{x}_o \\ \mathbf{y}_o \end{pmatrix}$$

erreicht. Anstelle der Inputreduktion ζ kann auch der Anteil des Inputvektors betrachtet werden, der nach der Reduktion der Inputs verbleibt. Dieser Anteil wird durch die Variable

$$\theta := 1-\zeta$$

ausgedrückt, sodass für die Referenz gilt:

$$\hat{\mathbf{z}} = \begin{pmatrix} -\theta\cdot\mathbf{x}_o \\ \mathbf{y}_o \end{pmatrix}.$$

Dabei ist die Maximierung von ζ äquivalent mit der Minimierung von θ:

$$\boxed{\begin{array}{l} \max \zeta \\ \zeta = \min\left\{\frac{1}{x_{o1}}\cdot s_1^-,\ldots,\frac{1}{x_{oM}}\cdot s_M^-\right\}\end{array}}$$

$$\Leftrightarrow \quad \boxed{\begin{array}{l} \min \theta \\ \theta = \max\left\{\left(1-\frac{1}{x_{o1}}\cdot s_1^-\right),\ldots,\left(1-\frac{1}{x_{oM}}\cdot s_M^-\right)\right\}\end{array}}$$

Diese Darstellungsweise der DEA-Zielfunktion ist allerdings nicht gebräuchlich. Daher wird sie durch die folgenden Transformationen im allgemeinen DEA-Modell in die übliche Gestalt überführt. Die Abweichungsvariablen in der Zielfunktion werden durch die Nebenbedingungen (1) aus (DEA), S. 39, umgestellt zu $s_m^- = x_{om} - \hat{x}_m$, ersetzt. Somit ist

$$\begin{aligned} \theta &= \max\left\{\left(1-\frac{1}{x_{o1}}\cdot(x_{o1}-\hat{x}_1)\right),\ldots,\left(1-\frac{1}{x_{oM}}\cdot(x_{oM}-\hat{x}_M)\right)\right\} \\ &= \max\left\{\frac{\hat{x}_1}{x_{o1}},\ldots,\frac{\hat{x}_M}{x_{oM}}\right\}. \end{aligned}$$

Für das zu minimierende Maß θ gilt daher

$$\theta \geqq \frac{\hat{x}_m}{x_{om}} \qquad \Leftrightarrow \qquad \hat{x}_m \leqq \theta \cdot x_{om}, \quad m = 1, ..., M,$$

was in Form von Nebenbedingungen im DEA-Modell berücksichtigt wird. Das Modell lautet somit

$$\begin{array}{rll}
(DEA^I) & & \\
\min & \theta & \\
\text{u.d.N.} & & \\
(1^*) & \hat{x}_m \leqq \theta \cdot x_{om}, & m = 1, ..., M \\
(2) & \hat{y}_n - s_n^+ = y_{on}, & n = 1, ..., N \\
(3) & s_n^+ \geqq 0, & n = 1, ..., N \\
 & \theta \in \mathbb{R} & \\
(4) & (-\hat{x}_1, ..., -\hat{x}_M, \hat{y}_1, ..., \hat{y}_N)^T \in TM. &
\end{array}$$

Der optimale Zielfunktionswert θ^* entspricht dem Debreu/Farrell-Maß. Er gibt an, ob die Produktion $\mathbf{z}_o$ inputorientiert radial effizient ist gemäß Definition 2.6, S. 23. Die Zielfunktion ist monoton steigend, sodass die gewählte Referenz schwach effizient ist. Um eindeutig eine effiziente Referenz zu erhalten, wird ein Effizienzmaß auf Basis einer ε-Maximin Metrik verwendet. Dazu wird der Zielfunktionswert um die mit der hinreichend kleinen Zahl ε gewichtete Summe der Abweichungen korrigiert:[127]

$$\begin{array}{rll}
(DEA_\varepsilon^I) & & \\
\min & \theta - \varepsilon \cdot \left(\sum\limits_{m=1}^{M} s_m^- + \sum\limits_{n=1}^{N} s_n^+ \right) & \\
\text{u.d.N.} & & \\
(1^{**}) & \hat{x}_m + s_m^- = \theta \cdot x_{om}, & m = 1, ..., M \\
(2) & \hat{y}_n - s_n^+ = y_{on}, & n = 1, ..., N \\
(3) & s_m^-, s_n^+ \geqq 0, & m = 1, ..., M, n = 1, ..., N \\
 & \theta \in \mathbb{R} & \\
(4) & (-\hat{x}_1, ..., -\hat{x}_M, \hat{y}_1, ..., \hat{y}_N)^T \in TM. &
\end{array}$$

Es sei θ^*, λ_j^*, $j = 1, ..., J$, s_m^{-*}, $m = 1, ..., M$, s_n^{+*}, $n = 1, ..., N$, die optimale

[127] Im Gegensatz zu der in Abbildung 2.8 vorgestellten ε-Maximin-Norm ist der Korrekturterm zu subtrahieren, da die Zielfunktion nun nicht mehr maximiert sondern minimiert wird.

Lösung von (DEA^I_ε). Die Produktion $\mathbf{z}_o$ wird mit der Referenz

$$\hat{\mathbf{z}} = \begin{pmatrix} -\theta^* \cdot \mathbf{x}_o - \mathbf{s}^{-*} \\ \mathbf{y}_o + \mathbf{s}^{+*} \end{pmatrix}$$

verglichen, wobei $\mathbf{s}^{-*} = (s_1^{-*}, ..., s_M^{-*})^T$ bzw. $\mathbf{s}^{+*} = (s_1^{+*}, ..., s_N^{+*})^T$ der Vektor der Input- bzw. Outputabweichungen im Optimum ist.[128] Für $\theta^* = 1$ gilt, dass die Produktion $\mathbf{z}_o$ schwach effizient ist. Eine weitere proportionale Inputreduktion ist innerhalb der Technologiemenge nicht möglich. Für die Überprüfung auf Pareto-Koopmans-Effizienz ist der DEA-Zielfunktionswert entscheidend. Ist dieser gleich eins, kann die Produktion $\mathbf{z}_o$ als effizient bezeichnet werden. In diesem Fall sind die optimalen Werte der Abweichungsvariablen null. Für positive Abweichungsvariablen ist der Zielfunktionswert kleiner als eins, sodass die Produktion $\mathbf{z}_o$ als nicht effizient beurteilt wird. Je kleiner der optimale Zielfunktionswert ist, desto größer ist die Abweichung von der effizienten Referenz und desto schlechter wird die DMU_o beurteilt. Auf Basis der Zielfunktionswerte aller DMUs ist eine Rangbildung möglich.

Der Effizienzwert θ wurde als im Vorzeichen unbeschränkte Variable $\theta \in \mathbb{R}$ eingeführt. Die Nebenbedingungen führen jedoch zu einer Beschränkung dieser Variablen.[129] Werden für die DMU_o semipositive Outputdaten angenommen ($y_{on} > 0$ für mindestens ein n), ist in den Nebenbedingungen (2) auch $\lambda_j > 0$ für mindestens ein j, sodass in den Nebenbedingungen (1) die Inputquantitäten der Referenz auch semipositiv sein müssen ($\hat{x}_m > 0$ für mindestens ein m).[130] Daraus folgt, dass $\theta > 0$. Des Weiteren gilt $\theta \leqq 1$, da das Modell eine zulässige Lösung bei $\theta = 1$, $\lambda_o = 1$, $\lambda_j = 0$ $(\forall j \neq o)$ hat. In diesem Fall entspricht die Produktion $\mathbf{z}_o$ ihrer schwach effizienten Referenz ($\mathbf{z}_o = \hat{\mathbf{z}}$). Somit ist θ faktisch auf das Intervall $0 < \theta \leqq 1$ beschränkt.

Zwar ist die Variable θ skaleninvariant, nicht jedoch die Abweichungsvariablen, die in der DEA-Zielfunktion als Korrekturwert enthalten sind.[131] Dies führt dazu, dass der korrigierte Zielfunktionswert ebenso wie die effiziente Referenz nicht skaleninvariant sind. Da ε sehr klein ist, ist diese Skaleninvarianz vernachlässigbar.

Die Abweichungsvariablen der Outputs s_n^+ entsprechen den absoluten Abwei-

[128] Vgl. u. a. ALI/SEIFORD 1993, S. 139.
[129] Vgl. auch im Weiteren COOPER et al. 2006, S. 43f.
[130] Sonst wäre auch die Annahme der Unmöglichkeit des Scharaffenlandes verletzt.
[131] Vgl. auch im Weiteren ALI/SEIFORD 1993, S. 144ff.

chungen Δy_{on} der Produktion $\mathbf{z}_o$ von der Referenz $\hat{\mathbf{z}}$, sodass gilt[132]

$$\Delta y_{on} \quad := \quad \hat{y}_n - y_{on} = s_n^{+*}, \quad n = 1, ..., N,$$

Die Abweichungsvariablen der Inputs erfassen - anders als im Modell (DEA), S. 39, - die über die proportionale Inputreduktion θ hinausgehenden Abweichungen, d. h.

$$s_m^{-*} \quad = \quad \theta^* \cdot x_{om} - \hat{x}, \quad m = 1, ..., M.$$

Die absoluten Abweichungen der Inputs können wie folgt berechnet werden:

$$\begin{aligned} \Delta x_{om} &:= x_{om} - \hat{x}_m \\ &= x_{om} - (\theta^* \cdot x_{om} - s_m^{-*}) \\ &= (1 - \theta^*) \cdot x_{om} + s_m^{-*}, \quad m = 1, ..., M. \end{aligned}$$

In der Literatur ist neben dem hier vorgestellten DEA-Modell mit der ε-Korrektur in der Zielfunktion auch ein 2-Phasen-Modell gebräuchlich.[133] Dabei wird in einem ersten Modell die Effizienz θ^* bestimmt und im zweiten Modell die Summe der Abweichungsvariablen maximiert. Bei dieser Vorgehensweise entfällt die Notwendigkeit, für die hinreichend kleine Zahl ε einen Wert festzulegen.[134] Das hier vorgestellte Modell fasst beide Phasen zusammen.

Das Modell (DEA_ε^I) ist noch bezüglich der Technologiemenge zu konkretisieren. Es entspricht unter der Annahme konstanter Skalenerträge (TM^{CRS}) dem originären DEA-Modell, das von Charnes, Cooper und Rhodes 1978 wurde

[132] Vgl. auch im Weiteren Cooper et al. 2006, S. 47.

[133] Zum 2-Phasen-Model vgl. u. a. Ali/Seiford 1993, S. 138ff; Cooper et al. 2006, S. 44f; Cooper et al. 2011b, S. 9ff.

[134] Vgl. Cooper et al. 2011b, S. 10f.

und nach diesen auch als *CCR-Modell* bezeichnet wird:[135]

$$
\begin{array}{lll}
(CCR) & & \\
\min & \theta - \varepsilon \cdot \left(\sum\limits_{m=1}^{M} s_m^- + \sum\limits_{n=1}^{N} s_n^+ \right) & \\
\text{u.d.N.} & & \\
(1) & \sum\limits_{j=1}^{J} \lambda_j \cdot x_{mj} + s_m^- = \theta \cdot x_{om}, & m = 1,\dots,M \\
(2) & \sum\limits_{j=1}^{J} \lambda_j \cdot y_{nj} - s_n^+ = y_{on}, & n = 1,\dots,N \\
(3) & \lambda_j, s_m^-, s_n^+ \geqq 0, & j = 1,\dots,J,\ m = 1,\dots,M,\ n = 1,\dots,N \\
 & \theta \in \mathbb{R}. &
\end{array}
$$

Die Interpretation einer optimalen Lösung entspricht der einer optimalen Lösung von (DEA_ε^I), S. 46.

Diese Formulierung des DEA-Modells wird in der Literatur als *Envelopment Form* bezeichnet, da die Konstruktion der Technologiemenge durch eine Umhüllung der beobachteten Produktionen erfolgt. Verwendet wird auch das zu (CCR) duale lineare Programm, die so genannte *Multiplier Form*. Es seien u_m, $m = 1,\dots,M$ und v_n, $n = 1,\dots,N$ die Dualvariablen der M Nebenbedingungen (1) und der N Nebenbedingungen (2), so ist das zu (CCR) duale lineare Programm:

$$
\begin{array}{lll}
(CCR_{Dual}) & & \\
\max & \sum\limits_{n=1}^{N} v_n \cdot y_{on} & \\
\text{u.d.N.} & & \\
(1) & \sum\limits_{m=1}^{M} u_m \cdot x_{om} = 1 & \\
(2) & \sum\limits_{n=1}^{N} v_n \cdot y_{jn} - \sum\limits_{m=1}^{M} u_m \cdot x_{jm} \leqq 0, & j = 1,\dots,J \\
(3) & v_n, u_m \geqq \varepsilon, & n = 1,\dots,N,\ m = 1,\dots,M.
\end{array}
$$

Bei dieser Modellformulierung werden die Variablen als Gewichte der Inputs und Outputs interpretiert. Das Modell (CCR_{Dual}) maximiert die Summe der bewerteten Outputs der DMU_o, wobei der Wert aller von DMU_o eingesetzten Inputs auf eins normiert wird (Nebenbedingung (1)). Der Wert der Outputs ist nicht größer als eins aufgrund der o-ten Bedingung der Nebenbedingungen (2) in Zusammenhang mit der Nebenbedingung (1). Zulässig sind diejenigen Input-Output-

[135] Vgl. CHARNES et al. 1978, S. 442.

Gewichte, bei welchen für alle DMUs der Wert ihrer Inputs nicht größer als der Wert ihrer Outputs ist (Nebenbedingungen (2)). Die Nebenbedingungen (2) und (3) stellen die duale Menge zur Technologie TM^{CRS} dar, die in Abschnitt 2.5.2 als P^{CRS} eingeführt wurde (Definition (2.4), S. 36). Aufgrund der Zielfunktionskorrektur im Modell (CCR) sind die Dualvariablen nun positiv (Nebenbedingung (3)), wobei ε eine hinreichend kleine Zahl ist.

Es sei v_n^*, $n = 1,...,N$, u_m^*, $m = 1,...,M$, die optimale Lösung von (CCR_{Dual}). Die Produktion $\mathbf{z}_o$ ist effizient, wenn der optimale Zielfunktionswert gleich eins ist. Je kleiner der Zielfunktionswert, desto schlechter wird die Produktion beurteilt. Die Variablen v_n^*, $n = 1,...,N$, und u_m^*, $m = 1,...,M$, stellen optimale Input- und Outputgewichte für Produktion $\mathbf{z}_o$ dar, d. h. es gibt keine andere Gewichtung, bei der die DMU_o unter der gegebenen Technologiemenge besser beurteilt wird.

Ausgehend vom CCR-Modell wurde von Banker, Charnes und Cooper 1984 ein DEA-Modell mit variablen Skalenerträgen formuliert.[136] Das nach ihnen benannte *BCC-Modell* unterscheidet sich vom CCR-Modell lediglich durch eine zusätzliche Nebenbedingung (Nebenbedingung (3)):

$$
\begin{array}{lll}
(BCC) & & \\
\min & \theta - \varepsilon \cdot \left(\sum\limits_{m=1}^{M} s_m^- + \sum\limits_{n=1}^{N} s_n^+ \right) & \\
\text{u.d.N.} & & \\
(1) & \sum\limits_{j=1}^{J} \lambda_j x_{mj} + s_m^- = \theta \cdot x_{om}, & m = 1,...,M \\
(2) & \sum\limits_{j=1}^{J} \lambda_j y_{nj} - s_n^+ = y_{on}, & n = 1,...,N \\
(3) & \sum\limits_{j=1}^{J} \lambda_j = 1 & \\
(4) & \lambda_j, s_m^-, s_n^+ \geqq 0, & j = 1,...,J, m = 1,...,M, n = 1,...,N \\
 & \theta \in \mathbb{R}. &
\end{array}
$$

Die Interpretation einer optimalen Lösung entspricht der einer optimalen Lösung von (DEA_ε^I), S. 46.

Unter Verwendung der bereits eingeführten Dualvariablen u_m, $m = 1,...,M$, und v_n, $n = 1,...,N$, sowie der Dualvariablen w für die zusätzliche Nebenbedin-

[136] Vgl. Banker et al. 1984, S. 1084ff.

gung (3), lässt sich das zu (BCC) duale Modell darstellen als

$$
\begin{array}{rll}
(BCC_{Dual}) & & \\
\max & \sum\limits_{n=1}^{N} v \cdot y_{jn} + \omega & \\
\text{u.d.N.} & & \\
(1) & \sum\limits_{m=1}^{M} u_m \cdot x_{om} = 1 & \\
(2) & \sum\limits_{n=1}^{N} v_n \cdot y_{jn} - \sum\limits_{m=1}^{M} u_m \cdot x_{jm} + \omega \leqq 0 & j = 1, ..., J \\
(3) & v_n, u_m \geqq \varepsilon, & n = 1, ..., N, m = 1, ..., M \\
 & w \in \mathbb{R}. &
\end{array}
$$

Da die Technologiemenge durch eine zusätzliche Nebenbedingung eingeschränkt wird und somit eine Teilmenge der im CCR-Modell verwendeten Technologiemenge darstellt ($TM^{VRS} \subseteq TM^{CRS}$), ist der optimale Zielfunktionswert des BCC-Modells mindestens genauso groß wie im CCR-Modell ($\theta^*_{BCC} \geqq \theta^*_{CCR}$).[137] Die Effizienz der DMU_o wird im BCC-Modell folglich gleich oder höher bewertet als im CCR-Modell.

[137] Vgl. COOPER et al. 2006, S. 88.

http://www.springer.com/978-3-658-14321-3

Effizienzanalyse von Dienstleistungsproduktionen
Eine Data Envelopment Analysis unter Berücksichtigung stochastischer externer Faktoren
Schlindwein, R.
2016, XX, 267 S. 53 Abb., Softcover
ISBN: 978-3-658-14321-3