56 ifo Forschungsberichte

Der Beitrag des öffentlichen Sektors zur Wertschöpfung: Messprobleme und Lösungsansätze

Thiess Büttner
Robert Fenge
Oliver Röhn
Thomas Strobel

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Forschungsbereich: Öffentlicher Sektor

Der Beitrag des öffentlichen Sektors zur Wertschöpfung: Messprobleme und Lösungsansätze

von

Thiess Büttner
Robert Fenge
Oliver Röhn
Thomas Strobel

München, Mai 2012

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Forschungsbereich: Öffentlicher Sektor

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über
http://dnb.d-nb.de
abrufbar.

ISBN: 978-3-88512-523-5

Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten.
Ohne ausdrückliche Genehmigung des Verlags ist es auch nicht gestattet, dieses
Buch oder Teile daraus auf fotomechanischem Wege (Fotokopie, Mikrokopie)
oder auf andere Art zu vervielfältigen.
© ifo Institut, München (2012)
Druck: ifo Institut, München

ifo Institut im Internet:
http://www.cesifo-group.de

Inhaltsverzeichnis

1. Einleitung 3

2. Outputmessung 8

2.1 Einleitung 8

2.2 Konzept: Produktionsfunktion 9

2.3 Konzept: Produktivität 10

2.4 Ausgangspunkt: Nominaler Output, Preise, Volumen, Mengen und Qualität 12

2.5 Aggregation heterogener Produkte mittels Indexzahlen 14
2.5.1 Axiomatischer und ökonomischer Ansatz der Indexbildung 14
2.5.2 Gewichte 18
2.5.3 Qualitätsadjustierung 21

2.6 Schlussfolgerung 22

3. Probleme und Lösungsansätze der Outputmessung im Nicht-Markt Sektor 23

3.1 Einleitung und Problemstellung 23

3.2 Kollektive vs. individuelle Güter 25

3.3 Verfahren der Outputvolumenmessung 27
3.3.1 Inputverfahren 27
3.3.2 Direkte Outputvolumenmessung 28
Aktivitäten 29
Outcome 30
Output 31

3.4 Aggregation heterogener Gruppen 32
Hypothetische Preise 33
Outcome basierte Gewichtung 33
Kostengewichte 34

3.5 Qualitätsmessung 36

3.6 Schlussfolgerungen 38

4. Statistische Übersicht zum öffentlichen Dienst im internationalen Vergleich 39

5. Illustration der Outputmessung anhand des Bildungssektors 48

5.1 Einleitung 48

5.2 Internationaler Methodenüberblick der Konstruktion von Output-Indizes 49
Deutschland 49
Finnland 51
Italien 51
Schweden 52
Vereinigtes Königreich 53

5.3 Simulationen von Outputindizes des Bildungssektors 58
Ungewichtete und kostengewichtete Outputindizes 59
Explizite Qualitätsadjustierung in den Outputindizes 61

5.4 Simulationen zur Produktivitätsentwicklung 64

6. Schlussfolgerung 66

Referenzen 68

Anhang 70

A1. Ober- und Untergrenzen des theoretischen Preisindex 70

A2. Internationale Output-Indizes 72

1. Einleitung

Gegenstand dieser Studie ist ein Vergleich und eine Bewertung von Methoden zur Messung der Produktivität des staatlichen Leistungsangebotes, also von Leistungen die nicht über einen Markt bereitgestellt werden. Die Produktivität gibt das Verhältnis der Ausbringungsmenge (Output) zur Menge der eingesetzten Produktionsfaktoren (Inputs) an. Output wie Input werden im Allgemeinen dabei zu Marktpreisen bewertet, um zu eindimensionalen Einheiten zu gelangen, die die Zusammenfassung unterschiedlicher Güter und Dienstleistungen, die sich im Zeitablauf auch ändern kann, erlaubt.

Das fundamentale Problem einer Messung staatlicher, nicht-marktbestimmter Dienstleistungen besteht darin, dass zur Bewertung dieser Dienstleistungen keine Marktpreise zur Verfügung stehen. Deshalb muss hier auf geeignete Mengenindikatoren für den Output zurückgegriffen werden, die hinreichend repräsentativ für die Erstellung der staatlichen Dienstleistungen sind und für die eine Gewichtung entwickelt wird, um sie zusammenfassen zu können. Zur Messung der Inputs werden die Kosten der Bereitstellung (z.B. Anzahl der Arbeitstunden bewertet mit dem Stundenlohn oder Abschreibungen auf den Kapitalstock) berechnet.

Kennzahlen der Produktivität spielen eine zentrale Rolle bei der Bewertung der gesamtwirtschaftlichen Entwicklung und der Entwicklung einzelner Branchen. So bemisst der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung regelmäßig auf der Basis der Produktivitätsentwicklung den lohnpolitischen Verteilungsspielraum und suggeriert damit, dass Lohn- und Gehaltsabschlüsse in dieser Höhe insgesamt ohne Beschäftigungseinbußen sind. Dabei wird in der Regel übersehen, dass die Produktivitätsentwicklung im öffentlichen Sektor eben nicht ohne weiteres bekannt ist. Vielmehr wird vom statistischen Bundesamt hier mit einer angenommenen Entwicklung im öffentlichen Sektor gerechnet (s.u.), die vergleichsweise geringe Steigerungen im Output nahe legt. Die Annahmen liegen deutlich unter den Durchschnittswerten in anderen Branchen, was aber mit der traditionellen Sicht vereinbar ist, wonach im Dienstleistungsbereich nur sehr geringe Produktivitätszuwächse zu verzeichnen sind.[1] Allerdings gibt es gerade in jüngerer Zeit aufgrund des zunehmenden EDV Einsatzes über den Produktivitätsfortschritt im Dienstleistungsbereich unterschiedliche Auf-

[1] Diese Sichtweise hat als die Baumol'sche Kostenkrankheit Eingang in die Literatur gefunden. Vgl. William J. Baumol, Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis, American Economic Review, 57, 3 (1967), S. 415-426.

fassungen.[2] Dies könnte bedeuten, dass der lohnpolitische Verteilungsspielraum insgesamt zu niedrig angesetzt ist.

Bedeutsamer sind aber die Konsequenzen fehlender Informationen über die Produktivitätsentwicklung im öffentlichen Sektor für die Lohn- und Personalpolitik im öffentlichen Sektor selbst. Ein verlässliches Maß für die staatliche Produktivität ist bedeutsam, weil es Aufschluss über die Funktionsfähigkeit des öffentlichen Sektors gibt, die im Interesse der Allgemeinheit liegt. Ohne Informationen über die Produktivität ist es für jede Regierung schwierig, öffentliche Ausgaben zu rechtfertigen, selbst wenn sie vielleicht für das Gemeinwesen sehr förderlich sind. Umgekehrt werden Kürzungen beim Mitteleinsatz als Einsparungen gewertet, auch wenn sie vielleicht das Leistungsangebot schwer beeinträchtigen.

Ziel dieser Untersuchung ist es vor diesem Hintergrund, die Problematik der Messung der Leistung des öffentlichen Sektors zu untersuchen und Ansätze für eine verbesserte Methode zur Bestimmung der Leistungsfähigkeit des öffentlichen Sektors zu diskutieren. Beispielhaft werden dabei verschiedene Varianten der Outputmessung auf den Bereich des Bildungssektors angewandt und die Unterschiede der Simulationsergebnisse diskutiert (siehe Abschnitte 5.3 und 5.4).

Grundsätzlich gibt es zwei Methoden zur Volumenmessung des Outputs im Bereich der Nichtmarkt-Dienstleistungen: die Input-Methode und die Output-Methode. Das Statistische Bundesamt in Deutschland hat in der Vergangenheit das Produktionsvolumen der Nichtmarkt-Dienstleistungen des Staates ausschließlich inputseitig, d.h. über die Berechnung der Volumenentwicklung der Kostenelemente (Vorleistungen, Abschreibungen, Arbeitnehmerentgelte und Nettoproduktionssteuern) bestimmt. Die preisbereinigte Entwicklung der Kostensumme dieser Inputs wurde mit dem Volumen des Outputs gleichgesetzt. Das Volumen der Inputs wurde dabei durch Deflationierung der einzelnen Kosten mit geeigneten Preisindizes ermittelt.

Der Nachteil der Input-Methode besteht darin, dass lediglich eine input-seitige Messung des Outputvolumens erfolgt. Eine Berechnung der Produktivität ist deshalb mit dieser Methode nicht möglich, da ja gerade der Faktor berechnet werden soll, um den die Ausbringungsmenge Output größer ist als die Einsatzmenge der Inputs. Das Statistische Bundesamt hat deshalb bei der Volumenberechnung im Bereich der nicht-marktbestimmten staatlichen Dienstleistungen

[2] Z.B. Jack E. Triplett, Barry P. Bosworth, 2003, Productivity measurement issues in services industries: "Baumol's Disease" has been cured, Economic Policy Review - Federal Reserve Bank of New York, Sep 2003.

die Annahme eines jährlichen Produktivitätsanstiegs von 0,5% zugrunde gelegt (vgl. Abschnitt 3.3.1).

Da die Volumenberechnungen des Outputs auch im Bereich staatlicher Dienstleistungen grundsätzlich für Produktivitätsuntersuchungen tauglich sein sollten, hat die Europäische Kommission im Jahr 2002 beschlossen, die direkte Volumenmessung auf Basis von Outputindikatoren in den nationalen Statistiken EU-weit vorzuschreiben. Diese so genannte Outputmethode greift auf Mengenindikatoren für einzelne messbare Dienstleistungen zurück und aggregiert dann den so gemessenen Output in den einzelnen Bereichen durch geeignet gewählte Gewichte. In einem Pilotprojekt in Deutschland wurde diese EU-Vorgabe für die drei Bereiche Bildung, Pflege und Interessenvertretung umgesetzt.

Ein Problem der Output-Methode wird darin gesehen, dass sie für individuell zurechenbare Dienstleistungen geeigneter ist, also solche staatlichen Dienstleistungen, die unmittelbar für einzelne Nachfrager bereitgestellt werden (z.B. Erziehung und Bildung, Kultur). Bei kollektiven Dienstleistungen, die für die Allgemeinheit und nicht für einzelne Haushalte erbracht werden (etwa Verteidigung, allgemeine Verwaltung), ist eine individuelle Transaktion zwischen Staat und Konsument nicht beobachtbar. Daher stellt sich die Schwierigkeit dar, geeignete, ausreichend repräsentative Outputindikatoren zu finden. Deshalb hat die Europäische Kommission entschieden, dass die Input-Methode weiterhin im Bereich dieser kollektiven Dienstleistungen zulässig sein soll (vgl. Abschnitte 3.1 und 3.2).

In Kapitel 2 dieser Studie wird zunächst die Methode der Outputmessung dargestellt, wie sie in der volkswirtschaftlichen Gesamtrechnung zur Ermittlung des Outputs privater Güter und Dienstleistungen, die auf einem Markt gehandelt werden, Anwendung findet. Dazu werden die grundlegenden ökonomischen Konzepte der Produktionsfunktion und der Produktivität erläutert und die auf Marktpreisen basierenden Indices dargestellt, die zur Aggregation der marktvermittelten Güter und Dienstleistungen herangezogen werden. Ein wichtiges Ziel dieses Kapitels ist es, die Bedeutung von Marktpreisen herauszuarbeiten, die unter anderem auch zur Qualitätsadjustierung des Outputvolumens wichtig sind.

Ausgehend von dieser theoretischen Sicht auf die Outputmessung von Marktgütern und Dienstleistungen werden in Kapitel 3 die Schwierigkeiten einer Anwendung dieser Messmethode bei staatlichen Dienstleistungen dargelegt. Aufgrund des Fehlens von Marktpreisen in diesem Bereich werden alternative Messmethoden diskutiert. In Deutschland war dies bis zum

Jahr 2005 die herkömmliche Methode, die Kosten des Inputeinsatzes zu kalkulieren, die nach Deflationierung den realen Output messen sollten. Insbesondere die Unmöglichkeit der Produktivitätsmessung, bei dem das Outputvolumen gerade ins Verhältnis zum Inputeinsatz gesetzt wird, hat dazu geführt, dass andere direkte Methoden der Outputvolumenmessung angewendet werden. Dabei wird auf Indikatoren für diesen Output zurückgegriffen, die entweder die Produktionsaktivitäten oder den Produktnutzen der Konsumenten messen sollen. Die Aggregation verschiedener staatlicher Güter und Dienstleistungen erfolgt dann durch eine Gewichtung, die entweder Preise vergleichbarer marktvermittelter Güter und Dienstleistungen heranzieht oder Gewichte über den Konsumnutzen beziehungsweise die Kosten konstruiert. Auch bei diesen Methoden tritt wieder das Problem der Qualitätsmessung auf. Mögliche Lösungsansätze, wie etwa Qualitätsänderungen verschiedener Dienstleistungen über die Veränderungen der Zusammensetzung homogener Leistungsgruppen zu erfassen (Qualitätsstruktureffekte), werden am Ende dieses Kapitels diskutiert.

In Kapitel 4 werden im Rahmen eines internationalen Vergleichs verfügbare Statistiken vorgestellt, die traditionell zur Beurteilung des Umfangs an bereitgestellten Leistungen durch den Staat herangezogen werden. Diese Statistiken sind ausschließlich Input basiert und unterliegen damit den vorher dargelegten Schwächen. Insbesondere lassen sie keinerlei Aussagen über die Veränderung der Produktivität zu. Zusätzlich gibt es von einigen internationalen Organisationen in jüngster Zeit Bestrebungen diese Inputmaße durch umfragebasierte Performance-Indikatoren zu ergänzen. Exemplarisch stellen wir einen Indikator der Weltbank vor, welcher die Qualität und Effektivität des Staatsektors zu messen versucht, und diskutieren dessen Vor- und Nachteile. Dieser kritische Überblick macht deutlich, dass die üblichen Vergleichszahlen ohne aussagekräftige statistische Outputmessungen kaum Aufschluss über die Wertschöpfung im öffentlichen Sektor geben können.

Kapitel 5 zeigt am Beispiel des Bildungssektors zunächst, welche unterschiedlichen Verfahren der statistischen Outputmessung in verschiedenen Ländern verwendet werden. Vor diesem Hintergrund werden einige innovative Ansätze für Output-Indices vorgestellt. So werden in Finnland für die Outputmessung im universitären Bereich die erworbenen Kreditpunkte als Indikator herangezogen, der sowohl Mengen- als auch Qualitätsmerkmale abbildet. In Italien wird im schulischen Bereich die Anzahl der Schüler pro Klasse als Qualitätsindikator verwendet, wobei die tatsächlichen Schülerzahlen je nach Klassengröße in standardisierte Schülerzahlen umgerechnet werden, um mögliche Unterauslastungen oder Überfüllungskosten zu erfassen. Aufbauend auf diesen Ergebnissen, werden Simulationen für den Bildungssektor

durchgeführt und die Simulationsergebnisse für die Output- und die Produktivitätsmessung unter Verwendung verschiedener Outputindices für den Bildungssektor in Deutschland vorgestellt und erläutert. Die Resultate für andere Länder (Frankreich, Großbritannien, Italien und Schweden) werden im Anhang dargestellt. Für Deutschland ist besonders auffällig, dass der Übergang von Aktivitätsindikatoren zu Outcome-Indikatoren, die den Konsumnutzen erfassen, zu einer deutlich anderen Einschätzung der Outputentwicklung führen kann.

Abschließend wird auf das Problem des Datenaufwands aufmerksam gemacht. Die Verwendung ausgefeilter Methoden zur direkten Outputmessung setzt voraus, dass eine sehr detaillierte Datenerhebung der statistischen Ämter für die jeweiligen Bereiche öffentlicher Dienstleistungen erfolgt. Deshalb kommt es zunächst darauf an, die benötigten Daten für die einzelnen Dienstleistungsbereiche zu benennen und auf eine Erhebung dieser Daten zu drängen. Erfolgreiche Anstrengungen dieser Art sind z.B. in Großbritannien zu beobachten, wo der Atkinson-Report zu einer Umstrukturierung der Datenerhebung für den öffentlichen Sektor geführt hat.

2. Outputmessung

2.1 Einleitung

Die Bestimmung verlässlicher Outputgrößen für erbrachte Leistungen des öffentlichen Sektors stellt in der Praxis oft ein schwieriges Unterfangen dar, da gängige Verfahren der Outputmessung, wie sie bei marktbezogenen Dienstleistungen Anwendung finden, oft nicht durchgeführt werden können. Messungen der Inputgrößen stellen dagegen auch bei nichtmarktbezogenen Dienstleitungen kein größeres Problem dar, da alle notwendigen Informationen über Kosten und Mengen vorhanden sind. Messproblematiken in den Outputgrößen verursachen jedoch nicht nur Fehlbewertungen des Outputniveaus, sondern induzieren auch Verzerrungen in den auf Output-Input-Verhältnissen basierenden Produktivitätskennzahlen. Neben Auswirkungen auf makroökonomischer Ebene, wie Fehleinschätzungen der wirtschaftlichen Lage von Sektoren hinsichtlich Outputniveau und Produktivitätswachstum, haben verzerrte Outputmaße auch mikroökonomische Relevanz, u.a. wenn sie als Basis für Lohnverhandlungen und Pensionsansprüche heranzogen werden.

Darüber hinaus wird die Outputmessung, wie sie in Form der volkswirtschaftlichen Gesamtrechnung (VGR) vom Statistischen Bundesamt durchgeführt wird, als ein wichtiger Wohlfahrtsindikator betrachtet, der die Beiträge ökonomischer Aktivitäten zur sozialen Wohlfahrt eines Landes misst. Dies gilt sowohl für private, als auch öffentliche ökonomische Aktivitäten, wenn auch der Beitrag ersterer Aktivitäten deutlicher messbar ist. Der Grund dafür liegt darin, dass private ökonomische Aktivitäten auf Märkten mit Preisbildungsmechanismen stattfinden, die Werteinschätzungen der Konsumenten und Kostenkalküle von Produzenten zum Ausgleich bringen. Fehlen derartige Märkte und somit die notwendigen Preisinformation, wie in vielen Bereichen des öffentlichen Sektors, so können dessen Beiträge zur Wohlfahrt nur schwerlich quantifiziert werden. Grundsätzlich bleibt bei einer Wohlfahrtsmessung im Rahmen von mit der VGR ermittelten Outputdaten zu berücksichtigen, dass unabhängig von privaten oder öffentlich bereitgestellten Dienstleistungen, hier ausschließlich der Marktwert der jeweiligen Leistungen erfasst wird. Der Teil der Wohlfahrt, der über die Konsumentenrente erschlossen wird, bleibt außen vor. Aus diesem Grund können Output und Wohlfahrt nicht gleichgesetzt werden. Jedoch signalisieren *Veränderungen* des Outputs *Veränderungen*

in der Wohlfahrt. Neben messbaren Outputdaten gibt es aber auch nicht quantifizierbare Faktoren wie Frieden, schönes Wetter und technische Innovationen, die die Wohlfahrt eines Landes nachhaltig beeinflussen können.

Aufgrund seiner makroökonomischen Bedeutung und seiner Auswirkungen auf die Wohlfahrt, befasst sich die vorliegende Studie mit Ansätzen der Outputmessung, insbesondere aber mit der Messung des statistisch schwer quantifizierbaren Outputs des öffentlichen Sektors (Nichtmarkt-Dienstleistungen). Die Relevanz dieses Sektors wird offenkundig, wenn man sich vergegenwärtigt, dass jener im Jahre 2005 einen Beitrag von ungefähr 18%[3] zur gesamtwirtschaftlichen Bruttowertschöpfung geleistet hat. In den folgenden Abschnitten werden zunächst allgemeine Konzepte der Produktions- und Produktivitätstheorie, sowie die Unterscheidung von nominalem, realem Output und Preisen näher erläutert. Anschließend werden unterschiedliche methodische Ansätze zur Berechung geeigneter Indexzahlen beschrieben, wobei verstärkt die Rolle von Preisen bei der Konstruktion von Gewichten zur Aggregation heterogener Güter und Dienstleistungen beleuchtet wird. Darauf aufbauend werden im folgenden Kapitel 3 spezielle Verfahren der Output- und Qualitätsmessung zur Konstruktion geeigneter Outputindizes für den öffentlichen Sektor vorgestellt.

2.2 Konzept: Produktionsfunktion

Zur analytischen Betrachtung des Outputs einer Firma oder eines Industriezweiges können grundsätzlich zwei Modellierungsansätze verwendet werden: Sie stellen entweder auf die Angebotsseite, oder äquivalent dazu, auf die Nachfrageseite ab. Die Angebotseite modelliert den eigentlichen Produktionsprozess der Firmen/Industriezweige, während die Nachfrageseite sich auf den Konsum und die Investitionen der Konsumenten und des Staates konzentriert. Beide Seiten (aggregiertes Angebot und aggregierte Nachfrage) determinieren dann die Mengen und Preise simultan.

Ökonomische Analysen der Angebotsseite greifen auf die Spezifikation einer so genannten Produktionsfunktion zurück. Dabei repräsentiert die Produktionsfunktion ein Konzept, das einen funktionalen Zusammenhang zwischen dem produzierten Output (Y) und den dafür verwendeten Input- oder Einsatzfaktoren herstellt. Für das Output-Konzept des gesamtwirt-

[3] Der Anteil bezieht sich dabei auf die Bereiche öffentliche Verwaltung, Verteidigung und Sozialversicherung, Erziehung und Unterricht, sowie Gesundheits-, Veterinär- und Sozialwesen.

schaftlichen Produktionswerts, der alle Güter und Dienstleistungen umfasst, die von einer Wirtschaftseinheit *i* innerhalb einer bestimmten Periode unter Weiterverarbeitung der Vorleistungen produziert werden, lässt sich folgende neoklassische Produktionsfunktion definieren:

$$Y_i = F(K_i, L_i, M_i, A_i) \quad (1).$$

Dabei reflektieren die jeweiligen Parameter das Technologieniveau (*A*) und die unter technisch effizienter Produktion eingesetzten Einsatzfaktoren Arbeit (*L*), Kapital (*K*) und Vorleistungen (*M*).[4] In der Realität besitzt jede Firma zwar ihre eigene Produktionsfunktion, aber aufgrund von Datenrestriktionen und zu Zwecken der Vereinfachung wird oft die Existenz einer aggregierten Produktionsfunktion auf Industrie- bzw. Makroebene unterstellt. Dies impliziert, dass der aggregierte funktionale Produktionszusammenhang aller Firmen innerhalb der betrachteten Einheit zumindest approximativ identisch ist.

In der unter (1) dargestellten Form, in der der produzierte Output, *Y*, den gesamtwirtschaftlichen Produktionswert darstellt, wird nicht nur der von der Wirtschafteinheit eigens produzierte Output, sondern es werden auch die Vorprodukte anderer Einheiten berücksichtigt. Betrachtet man für *Y* stattdessen die Bruttowertschöpfung, d.h. den Produktionswert unter Ausschluss der Vorleistungen, so dass nur noch die Eigenleistung der betrachteten Wirtschafteinheit im Vordergrund steht, erhält man folgende Produktionsfunktion:

$$Y_i = F(K_i, L_i, A_i) \quad (2).$$

Die Existenz einer Produktionsfunktion für das Bruttowertschöpfungs-Konzept unterstellt implizit, dass der Produktionsprozess zur Erstellung der Bruttowertschöpfung separat vom gesamtwirtschaftlichen Produktionswert betrachtet werden kann. Der Vorteil eines solchen Konzepts liegt in der Vermeidung von potentiellen Doppelzählungen von Produkten innerhalb des eigentlich betrachteten Produktionsprozesses, die aus anderen Produktionsprozessen stammen und somit fälschlicherweise der Outputleistung der verarbeitenden Einheit zugerechnet werden.

2.3 Konzept: Produktivität

Bei der Berechnung von Produktivitätskennzahlen kann auf mehrere Ansätze zurückgegriffen werden, wobei Produktivität sich grundsätzlich als eine Form von Output-Input-Verhältnis

[4] Die Subskripte für die Zeitpunkte *t* wurden zur Vereinfachung weggelassen.

definieren lässt. Als Einteilungsmerkmal der zur Verfügung stehenden Konzepte kann die Anzahl der zugrunde liegenden Produktionsfaktoren herangezogen werden. Dabei wird meist eine Unterteilung in „Einzel-Faktor-Produktivitäts-Maße (single-factor productivity) oder „Multi-Faktor-Produktivitäts-Maße (multi-factor productivity) vorgenommen. Die Wahl des zu verwendenden Produktivitätskonzepts wird dabei vor dem Hintergrund des bei der Messung verfolgten Zwecks der Untersuchung sowie der Verfügbarkeit der Daten getroffen.

Wird der Output zu einem Input-Faktor (bspw. Arbeit *L*) in Verhältnis gesetzt, dann spricht man von einem „Einzel-Faktor-Produktivitäts-Maß, wie z.B. bei der Arbeitsproduktivität (AP). Diese betrachtet den produzierten Output zu eingesetzter Arbeit, wobei letztere häufig in Form von gearbeiteten Stunden oder der Anzahl der Beschäftigten gemessen wird:

$$AP = \frac{Y}{L} \tag{3}.$$

Stehen dagegen mehrere Produktionsfaktoren (d.h. *K*, *L*, und *M*) zur Verfügung, können an stelle einer Betrachtung von Einzel-, auch Multi-Faktor-Produktivitäts-Maße konstruiert werden. Die zuvor besprochene Produktionsfunktion liefert das Grundgerüst für diese Art von Produktivitätsmessungen, denn sie ermöglicht es den effizienten Einsatz kombinierter Input-Faktoren (über den funktionalen Zusammenhang in F) bei der Produktion des Outputs zu untersuchen. Sie bereitet dadurch das theoretische Fundament für die formale Herleitung der Multi-Faktor-Produktivität, die häufig auch als totale Faktorproduktivität (*TFP*) bezeichnet wird. Unter der Annahme einer neoklassischen Produktionsfunktion nach Gleichung (1) und Hicks-neutralem technologischen Fortschritt, *A*, lässt sich die totale Faktorproduktivität wie folgt definieren:[5]

$$Y = AF\left(K, L, M\right) \tag{4},$$

$$A = TFP = \frac{Y}{F\left(K, L, M\right)} \tag{5}.$$

Dies gilt analog auch für eine Produktionsfunktion nach dem Bruttowertschöpfungs-Konzept. Sollten sich die Input-Faktor neben der Unterscheidung in Kapital (*K*), Arbeit (*L*) und Vorleistungen (*M*), weiter in Energie (*E*) und Services (*S*) disaggregieren lassen, so können so genannte „KLEMS Multi-Faktor-Produktivitäts-Maße“ berechnet werden. Diese sind per de-

[5] Hicks-Neutralität impliziert, dass sich technologischer Fortschritt in Form einer multiplikativen Konstante definieren lässt, die bei konstanten Faktorpreisen zu einer Zunahme des Produktionspotentials *Y* führt, ohne dabei auf einen der Input-Faktoren im Speziellen verstärkt zu wirken.

finitionem nur für den Output-Fall des gesamtwirtschaftlichen Produktionswerts konstruierbar. Auch wenn die folgenden Ausführungen weitere Betrachtungen der Produktivitätsaspekte ausblenden, so zeigt sich doch bereits in den hier beschriebenen Konzepten die Bedeutung, die einer möglichst genauen Messung des Outputs zukommt. Jegliche Art von Effizienzmessung ist somit auf eine möglichst genaue Bestimmung des Outputs angewiesen.

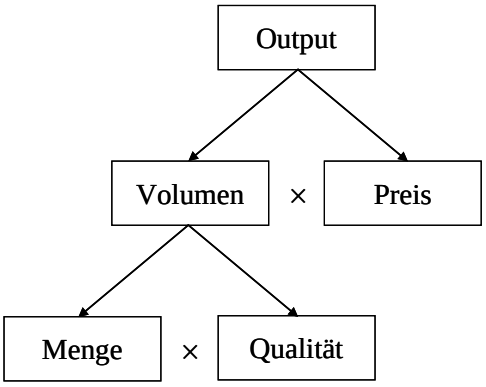
2.4 Ausgangspunkt: Nominaler Output, Preise, Volumen, Mengen und Qualität

Die Implementierung der theoretischen Outputkonzepte mit dem vorhandenen statistischen Datenmaterial verlangt nach einer genauen Unterscheidung in nominale und reale Outputgrößen. Der folgende Abschnitt beschäftigt sich deshalb ausschließlich mit der Messung der Output-Größe, *Y*, und ist für beide Konzepte (Produktionswert, Bruttowertschöpfung) gültig. Grundsätzlich lässt sich eine Beschreibung des Outputs im Fachterminus nach jeweiligen und konstanten Preisen durchführen. Eine Messung in jeweiligen Preisen (nominal) bezeichnet dabei den wertmäßigen Bestand des Outputs zu aktuellen Preisen der jeweiligen Berichtsperiode. Dahingegen wird bei der Bewertung in konstanten Preisen der Output einer bestimmten Berichtsperiode in Preisen einer anderen Periode ausgewiesen. Letztere ermöglicht es somit die Preisbewegung konstant zu halten und nur Veränderungen der Mengen (d.h. reale Veränderungen) in Preisen der Bezugsperiode zu betrachten. Diese Trennung von nominalem Output in eine Mengen- (auch Volumen genannt) und eine Preiskomponente, ermöglicht es, die Entwicklung der Inflation von der der produzierten Mengen zu isolieren (siehe schematische Abbildung 2.1). In der Praxis erfolgt die Messung der Preis- und Mengenkomponenten über die Verwendung von Indizes.

Im Idealfall bildet ein Preisindex ausschließlich die Preisentwicklung ab. Ist dies der Fall, dann haben beobachtbare Veränderungen im nominalen Output ihren Ursprung entweder in der Veränderung einer der beiden Komponenten (Preis-, Mengenindex) oder sie sind auf Veränderungen beider Komponenten zurückzuführen. Veränderungen in der Qualität bleiben bislang unberücksichtigt. Prinzipiell lässt sich das Bestreben einer Trennung in reale und nominale Output-Änderungen, neben dem vorherrschenden statistischen Zweck der preisunabhängigen Vergleichbarkeit von Outputbeständen, auch über die ökonomische Theorie der Klassischen Dichotomie begründen, in der langfristig nur die realen Entwicklungen von ökonomi-

schen Größen für das Entscheidungskalkül von Konsumenten und Produzenten relevant sind (Superneutralität des Geldes).[6]

Abb. 2.1: Schematische Darstellung der Outputmessung



Insofern es keine Veränderungen in der Qualität von Waren und Produkten gibt, bilden preisbereinigte Outputgrößen die Menge bzw. das Volumen ab. Verändert sich neben den Mengen zusätzlich die Qualität der Produkte, so lässt sich ein Volumenindex weiter aufspalten in zwei Subindizes, wovon der eine die reine Mengen- und der andere die Qualitätsveränderungen misst (siehe schematische Abb. 2.1). Diese weiterführende Unterteilung macht bereits die Problematik der Messung von Qualitätsveränderungen deutlich. Eigentlich sollen Veränderungen in der Produktqualität letztlich in der Volumen- und nicht in der Preiskomponente erfasst werden. Üblicherweise finden sich diese Effekt aber in der Preiskomponente wieder, da Veränderungen in der Qualität einer bestimmten Zahlungsbereitschaft der Konsumenten gegenüberstehen, die sich dann wiederum in Veränderungen der Preise äußern. Ein Außerachtlassen der Qualitätsadjustierung von Preisen hat demnach zur Folge, dass reale Größen aufgrund eines zu hoch angesetzten Preis-Deflators, nach der Deflationierung einen zu geringen realen Wert bzw. Wachstumsrate ausweisen werden. Dadurch entsteht dann ein verzerrtes Bild der tatsächlichen realen Output- und Produktivitätsentwicklung.

Alternativ zur Bestimmung des realen Output-Volumens mittels Preis-Deflationierung können direkte Outputmessungen durchgeführt werden. Dabei werden Mengen- und Qualitätsindizes von vornherein separat bestimmt und anschließend zur Konstruktion des Volumenindex zusammengeführt. Der Nachteil der direkten Methode besteht darin, dass zur Volumenbestim-

[6] Die Klassische Dichotomie ist ein zentraler Bestandteil der Klassischen und Neo-klassischen Makroökonomik, während sie von Keynesianern und Monetaristen angezweifelt wird. Letzteren zu Folge sind Preise in der kurzen Frist fix und passen sich somit nicht sofort an etwaige Schocks an, so dass Änderungen im Geldangebot die aggregierte Nachfrage und schließlich auch reale makroökonomische Größen beeinflussen können.

mung deutlich größere Stichproben der einzelnen Waren vorhanden sein müssen, um damit die nötige Repräsentativität zu gewährleisten. Dies ist darauf zurück zuführen, dass die einzelnen Waren sehr unterschiedliche Ausprägungen, je nach Art der Produktion, aufweisen können. Die Deflationierungsmethode bringt außerdem den Vorteil mit sich, dass Qualitätsadjustierungen aufgrund von fixen Warenkörben deutlich besser bewerkstelligt werden können.

2.5 Aggregation heterogener Produkte mittels Indexzahlen

Aufgrund der enormen Anzahl hergestellter Produkte, sowie den dazugehörigen Preisen, bedarf es eines Aggregationskonzepts, um diese zu einigen wenigen, kompakten Kennzahlen zur Darstellung des gesamtwirtschaftlichen Outputs einzelner Wirtschaftseinheiten (z.B. Industriezweige) zu verdichten. Diese Verdichtung von Informationen geschieht in Form von Indizes. Dabei stellt sich die Frage, wie die Vielzahl von mikroökonomischen Informationen unter Berücksichtigung der ökonomischen Theorie am besten in Indexzahlen aggregiert werden kann. Indexzahlen ermöglichen es, zeitliche Vergleiche aggregierter Veränderungen anzustellen, wobei sie eine dem Vergleich verschiedener Erscheinungen dienenden Bezugnahme auf eine Basis- oder Referenzperiode aufweisen. In den folgenden beiden Abschnitten soll auf die theoretischen Grundlagen der Indexbildung, sowie auf die Gewichtungsproblematik im Zuge der Aggregation heterogener Produkte näher eingegangen werden.

2.5.1 Axiomatischer und ökonomischer Ansatz der Indexbildung

Grundsätzlich bieten sich drei Ansätze zur Rechtfertigung der Bildung von Indexzahlen: der axiomatische, der ökonomische und der stochastische. Aufgrund ihrer in der Literatur dominanten Stellung wird im Folgenden nur auf die ersten beiden Ansätze näher eingegangen. Der axiomatische Ansatz betrachtet Preis- und Mengenkomponente als voneinander unabhängige Variable. Ein Index wird demnach als eine Funktion der beobachteten Preise und Mengen definiert, wobei diese Funktion bestimmte axiomatische Bedingungen bzw. Tests erfüllen muss.[7] Dabei geht es weniger um die Modellierung ökonomischer Erkenntnisse nach der Konsumenten- bzw. Produzententheorie, sondern stattdessen um die Wahl eines idealen Indexes, der möglichst viele der umfangreichen technischen Anforderungen besteht. Als problematisch werden diesbezüglich die fehlende inhaltliche (d.h. ökonomische) Interpretierbarkeit,

[7] Eine umfangreichere Übersicht der relevanten Axiome und Tests findet sich in Diewert (2007).

die praktische Durchführbarkeit von Indexberechnungen, sowie die Verständlichkeit der Indexaussage gesehen.

Der ökonomische Ansatz dagegen folgt der ökonomischen Theorie und unterstellt demgemäß eine Abhängigkeit der Mengen von den Preisen, d.h. $q = f(p)$. Diese Abhängigkeit beruht auf der Annahme, dass Konsumenten in kompetitiven Märkten als Preisnehmer agieren und die Preise nicht direkt beeinflussen können. Demnach sieht sich der Konsument den Preisen $p \equiv (p_1, \ldots, p_n)$ für *n* Güter $q \equiv (q_1, \ldots, q_n)$ gegenüber, wobei er versuchen wird, seine Kosten, bei gegebenem Nutzenlevel *u* und Präferenzen *F*, über die Wahl der optimalen Gütermenge q^* zu minimieren. Die Kostenfunktion *C* und das dazugehörige Minimierungsproblem des Konsumenten lassen sich dann wie folgt definieren:

$$\begin{aligned} C(p,u) &\equiv \min_{q^*} \{p \cdot q : F(q) \geq u\} \\ &= \sum_{i=1}^{n} p_i q_i^* \end{aligned} \qquad (6),$$

wobei $p \cdot q$ das innere Produkt der Preis- und Mengenvektoren, *p* und *q*, darstellt. Anhand Gleichung (6) wird deutlich, dass davon ausgegangen wird, dass die in der Realität beobachteten konsumierten Güter q_n die optimale Lösung q_n^* für das zugrunde gelegte Minimierungsproblem liefern (vgl. Theorie der bekundeten Präferenzen[8]). Im Rahmen der Konsumententheorie wählt der Konsument seine optimal nachgefragten Mengen immer so, dass der Grenznutzen einer zusätzlich konsumierten Einheit genau den dafür vorgegebenen Preisen entspricht. Aus diesem Grund kommt den Preisen eine entscheidende Bedeutung bei der Konstruktion von ökonomisch motivierten Indexzahlen zu. Diese reflektieren nämlich die marginalen Zahlungsbereitschaften der Konsumenten.

Der wahre zugrunde liegende Preisindex lässt sich demnach für den Zwei-Perioden-Fall, *t* = 0, 1, gemäß der Kostenfunktion in (6) als

$$P^t(p^0, p^1, q^*) \equiv \frac{C^t\left[p^1, F(q^*)\right]}{C^t\left[p^0, F(q^*)\right]} \qquad (7)$$

definieren. Bei gegebenem Nutzenniveau $u \equiv F(q)$ liefert $P^t(p^0, p^1, q^*)$ die minimalen Kosten in *t* für einen gegeben Preisvektor p^1, relativ zu den minimalen Kosten in *t* zur Erreichung

[8] Die Theorie der bekundeten Präferenz ist eine Methode, über die die bestmögliche Option eines Entscheidungskalküls auf Grundlage des in der Realität beobachteten Konsumentenverhaltens festgestellt werden kann, vgl. dazu Samuelson (1938).

desselben Nutzenniveaus u für den gegebenen Preisvektor p^0. Dabei ist wichtig, dass sich die Güterkomponenten über die Zeit hinweg nicht verändern.[9] Dies ist ein charakteristisches Kennzeichen für einen nach der ökonomischen Theorie definierten Preisindex, d.h. außer den Preisen werden alle sonstigen Umweltvariablen konstant gehalten. $P^t\left(p^0, p^1, q^*\right)$ wird nach der Literatur und aufgrund der verwendeten Kostenfunktionen auch als True Cost-of-Living Index oder wahrer Index der Lebenshaltungskosten bezeichnet (Konüs, 1924).

Da die Konsumentenpräferenzen und somit auch die Kostenfunktionen in der Realität jedoch unbekannt sind, ist die exakte Umsetzung von ökonomisch-theoretisch motivierten Indizes in der Praxis oft nicht möglich. Als möglichst genaue Approximation, die mittels beobachteter Daten erreicht werden kann, werden in der Regel Laspeyres- (P_L) (1871) und Paasche-Preisindizes (P_P) (1874) verwendet:

$$P_L\left(p^0, p^1, q^0, q^1\right) = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \tag{8},$$

$$P_P\left(p^0, p^1, q^0, q^1\right) = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} \tag{9}.$$

Dabei beschreiben beide Indizes die Preisentwicklungen der Waren $i = 1, \ldots, n$ zwischen den Perioden 0 und 1, d.h. p_i^0 und p_i^1, jeweils von einem unterschiedlichen Standpunkt aus betrachtet. Während nach P_L die Preisentwicklungen zwischen 0 auf 1 unter fixen Mengen in 0, q_i^0, berechnet werden, erfolgen bei P_P die Berechnungen der Preisveränderungen zu fixen Mengen in 1, q_i^1.

Obwohl beide Indizes zur Berechnung der Preis- und somit der Kostenveränderungen für sich genommen plausibel sind, betrachten sie die Preisveränderung dennoch von unterschiedlichen Standpunkten aus und führen in der Praxis meist zu unterschiedlichen Ergebnissen. So ermitteln beide Indizes die Preisentwicklungen in Bezug auf unterschiedliche Warenkörbe, wodurch Abweichungen in den Preisstatistiken auftreten können. Unter Annahme eines kostenminimierenden Konsumentenverhaltens kann gezeigt werden, dass beide Indizes die Ober- und Untergrenze für den wahren Index der Lebenshaltungskosten nach Gleichung (7) abbil-

[9] Im hier vorliegenden Fall der Gleichung (7) wird den Mengen kein konkreter Superskript zugewiesen. Der Superskript für *t* könnte sich sowohl auf *t*=0 oder *t*=1 beziehen, wird aber aufgrund der Intention, eine allgemeine Darstellung von Preisindizes zu liefern, zunächst außer Acht gelassen.

den (Konüs, 1924).[10] Da es jedoch erstrebenswert ist, eine einzelne Veränderungsrate zu berechnen und dabei den wahren Index möglichst genau zu approximieren, ist es erforderlich, mit einer Indexformel die beiden Standpunkte miteinander zu vereinigen. Diese Vereinigung wird durch die Indexformeln von Fisher (1922) und Thörnqvist (1936) gewährleistet, welche in der nachfolgenden Box kurz behandelt werden.

Die zuvor beschriebenen Preisindizes werden anschließend zur Deflationierung der nominalen Outputindizes verwendet. Die Bestimmung der Volumenindizes erfolgt demnach implizit, d.h. durch das Verhältnis von nominalen Output- zu Preisindizes. Als Ausgangspunkt lässt sich der nominale Outputindex, *W*, spezifizieren als:

$$W\left(p^0, p^1, q^0, q^1\right) = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} = P\left(p^0, p^1, q^0, q^1\right) \times Q\left(p^0, p^1, q^0, q^1\right) \quad (10),$$

mit den entsprechenden Preis- und Volumenindizes, *P* und *Q*. Die Auflösung nach *Q* ermöglicht dann die implizite Berechnung des Volumenindex.[11]

Betrachtet man bspw. den Laspeyres-Volumenindex, so lässt sich dieser wie folgt ermitteln:

$$\begin{aligned} Q_L\left(p^0, p^1, q^0, q^1\right) &= \frac{W\left(p^0, p^1, q^0, q^1\right)}{P_P\left(p^0, p^1, q^0, q^1\right)} \\ &= \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0 \cdot \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1}} \\ &= \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} \left(q_i^0 / q_i^0\right) p_i^0 q_i^1}{\sum_{j=1}^{n} p_j^0 q_j^0} \\ &= \sum_{i=1}^{n} s_i^0 \frac{q_i^1}{q_i^0} \end{aligned} \quad (11).$$

Die Mengenveränderungen reflektieren demnach einfach das gewichtete arithmetische Mittel der *n* Mengenverhältnisse q_i^1 / q_i^0, wobei die Gewichte die wertmäßigen Anteile der Periode 0 (z.B. Referenzperiode) darstellen:

[10] Dabei liefert der Laspeyres-Preisindex für den Fall eines Kostenminimierungsproblems die Ober-, der Paasche-Preisindex die Untergrenze. Für eine detailiertere Herleitung siehe Anhang A1.

[11] Bei der Berechung der Volumenindizes ist zu berücksichtigen, dass der reale Laspeyres-Volumenindex unter Verwendung des Paasche-Preisindizes gewonnen wird, und umgekehrt.

$$s_i^0 = \frac{p_i^0 q_i^0}{\sum_{j=1}^{n} p_j^0 q_j^0} \qquad (12).$$

2.5.2 Gewichte

Der Verwendung von entsprechenden Gewichten kommt bei der Aggregation über unterschiedliche Güter in der ökonomisch motivierten Indextheorie eine besondere Bedeutung zu. Würde man nur ein undifferenziertes Gut betrachten, so wären keine Gewichte notwendig. Bei der Gewichtung mehrer einzelner Güter, oder im hier vorliegenden Fall Output-Komponenten, zu einem Gesamtaggregat sollen die jeweiligen Güter/Einzelkomponenten gemäß ihrem ökonomischen Gewicht in das Aggregat eingehen. Zentraler Bestandteil der Gewichte sind gemäß der obigen Ausführung die Preise.

Grundsätzlich spiegeln Preise die marginalen Wertschätzungen der Marktteilnehmer der von ihnen konsumierten Güter wider und erfüllen somit mehrere Funktionen. Zum einen beinhalten sie eine Signalfunktion, da sie einen Indikator für die Knappheit von Gütern darstellen. Veränderungen in den Preisen signalisieren den Marktteilnehmern, dass sich Angebots- und Nachfragestrukturen für bestimmte Güter verändert haben. Gleichzeitig übernehmen sie eine Koordinierungsfunktion, wodurch die Marktteilnehmer ihre individuellen Präferenzen aufgrund der veränderten Marktlage neu aufeinander abstimmen können. Sie beinhalten aber auch eine Allokationsfunktion durch die Produktionsfaktoren, wie bspw. Arbeit und Kapital, an den Ort ihrer höchsten Effizienz gelenkt werden. Voraussetzung dafür, dass Preise die beschriebenen Funktionen erfüllen können, ist die Existenz eines funktionierenden Marktes, der gewährleistet, dass Konsumenten ihre marginalen Zahlungsbereitschaften über Preise ausdrücken können.

Wichtig im Hinblick auf die Verwendung von Preisen in den Indexgewichten ist, dass diese marginale Zahlungsbereitschaften reflektieren. Im Beispiel des Kostenminimierungsproblems der Konsumenten entsprechen die Preise den marginalen Zahlungsbereitschaften (Grenznutzen), die die Konsumenten bereit sind für den Konsum einer weiteren, zusätzlichen Gütereinheit auszugeben. Auf der Produzentenseite folgt bei kompetitiven Faktormärkten, dass die Grenzkosten der Güterproduktion ebenfalls den Preisen entsprechen. Folglich wären, unter diesen Bedingungen, die am Markt beobachteten Preise mit den Zahlungsbereitschaften der Konsumenten und den Grenzkosten der Produzenten identisch. Die Preise würden demzufol-

ge alle notwendigen Informationen für die Durchführung eines optimalen Tauschgeschäfts zwischen den Marktteilnehmern in sich vereinigen. In der Realität können Grenzkosten und Grenznutzen jedoch auseinander fallen.

Fisher-Index

Da Laspeyres und Paasche unterschiedliche Schätzungen für die Preisentwicklung generieren, ist eine Indexformel, die die Informationen beider Indizes miteinander verbindet, wünschenswert. Unter bestimmten Annahmen an die im Kostenminimierungskalkül der Konsumenten zugrunde gelegten Präferenzfunktionen kann gezeigt werden, dass der theoretische Preisindex invariant zu dem unterstellen Nutzenlevel ist. Der Laspeyres- und der Paasche-Index liefern somit die Ober- bzw. Untergrenze zu demselben theoretischen Preisindex. Aufgrund dieser Eigenschaft ollte die Bildung einer Art Durchschnitt aus beiden Indizes nach Möglichkeit die beste Approximation an den theoretischen Preisindex darstellen. Eine Approximation dieser Art wird bspw. durch den häufig verwendeten Fisher-Index (1922) bewerkstelligt. Dieser ergibt sich als der geometrische Durchschnitt aus Laspeyres- und Paasche-Index:

$$P_F\left(p^0, p^1, q^0, q^1\right) \equiv P_L\left(p^0, p^1, q^0, q^1\right)^{\frac{1}{2}} P_P\left(p^0, p^1, q^0, q^1\right)^{\frac{1}{2}} \qquad \text{(B1).}$$

Der Fisher-Index weist besondere Eigenschaften auf, die ihn zu einem bevorzugten Index machen. Er stellt eine Approximation von flexiblen Aggregator-Funktionen (z.B. Nutzen- und Produktionsfunktionen) dar, wobei die Approximation für bestimmte Aggregator-Funktionen exakt ist. Aufgrund dieser Tatsache zählt man ihn auch zu den so genannten superlativen Indizes (Diewert, 1976). Die am Beispiel von Preisindizes dargestellte Bildung eines Fisher-Index lässt sich aufgrund des für diesen Index gültigen Produkt-Tests auch in Form eines Volumenindex schreiben (Diewert, 1987).

Thörnqvist-Index

Da die Gültigkeit des Fisher-Index als Approximation des theoretischen Index an die grundlegende Annahme von homothetischen Präferenzen gebunden ist und diese im Falle von Nutzenfunktionen eher als unrealistisch betrachtet wird, kann stattdessen auf einen ebenfalls weit verbreiteten superlativen Index, den so genannten Thörnqvist-Index (1936), zurückgegriffen werden. Jener wird auch in den späteren Untersuchungen dieser Studie zur Anwendung kommen.

Der Thörnqvist-Index hat die besondere Eigenschaft, dass er Indizes, die für zeitlich kontinuierliche Dimensionen konstruiert werden (z.B. Divisia-Index, 1925), unter Verwendung von diskreten Beobachtungszeitpunkten approximiert. Dies ist von Vorteil, da in der Realität meistens auf diskrete Zeitpunkte (Quartals- oder Jahresdaten) zurückgegriffen werden muss, während Konsumenten- bzw. Produzentenverhalten mit längerem Zeithorizont in den Wirtschaftswissenschaften vermehrt unter einer stetigen Zeitdimension modelliert wird. Außerdem birgt er, wie bereits oben angesprochen, die weitere Eigenschaft in sich, dass er nicht auf die Annahme homothetischer Präferenzen angewiesen ist.

Der Thörnqvist-Index ist folgendermaßen definiert:

$$P_T\left(p^0, p^1, q^0, q^1\right) = \prod_{i=1}^{n} \left(\frac{p_i^1}{p_i^0}\right)^{\left(s_i^0 + s_i^1\right)/2} \qquad \text{(B2),}$$

wobei s_i wiederum die Gewichte in Form von wertmäßigen Anteilen der jeweiligen Periode 0 und 1 darstellen. Dieser Index besitzt neben seiner approximativen Eigenschaft an flexible Aggregator-Funktionen den weiteren Vorteil, dass er für homogene transzendent-logarithmische (translog) Aggregator-Funktion exakt ist. Diese Eigenschaft ist insbesondere aus dem Grund nützlich, da Aggregator-Funktionen vom Translog-Typ wiederum Aggregator-Funktionen anderen Typs approximieren. Somit ist eine Verwendung des Thörnqvist-Index auch dann sinnvoll, wenn die reale Welt keine translog Welt darstellt. Dennoch haben beide Indizes, der Fisher als auch der Thörnqvist, aufgrund ihrer superlativen Eigenschaft ähnliche theoretische Eigenschaften und sind approximativ gleich (Diewert, 1978). Durch Logarithmierung erhält man schließlich den zu verwenden Index:

$$\ln P_T\left(p^0, p^1, q^0, q^1\right) = \sum_{i=1}^{n} \frac{1}{2}\left(s_i^0 + s_i^1\right) \ln\left(\frac{p_i^1}{p_i^0}\right) \qquad \text{(B3).}$$

Aufgrund des auch hier gültigen Produkt-Tests lässt sich dieser Index analog ebenfalls in einen Volumenindex überführen (Diewert, 1978), wodurch Gleichung (B3) ebenso in Form von

$$\ln Q_T\left(p^0, p^1, q^0, q^1\right) = \sum_{i=1}^{n} \frac{1}{2}\left(s_i^0 + s_i^1\right) \ln\left(\frac{q_i^1}{q_i^0}\right) \qquad \text{(B4)}$$

geschrieben werden kann. Diese Notation wird dann im Rahmen der später konstruierten Output-Indizes relevant sein.

2.5.3 Qualitätsadjustierung

Die Hauptschwierigkeit bei der Bestimmung verlässlicher Outputindizes liegt in der Qualitätsadjustierung der realen Mengen. Zur Qualitätsadjustierung bedarf es einer Extraktion der Qualitätskomponente aus den Preisen. Die Trennung der Preisentwicklung in eine reine Preis- und eine Qualitätskomponente wurde bereits unter 1.4 angesprochen und basiert darauf, dass Veränderungen in beobachteten Güterpreisen sowohl einen Inflationseffekt als auch einen Effekt, der durch veränderte Warencharakteristika induziert wird, beinhalten. Änderungen der Warencharakteristika reflektieren Qualitätsveränderungen, die von den Konsumenten wiederum über ihre Zahlungsbereitschaften bewertet werden.

Die Messung des reinen Preiseffekts (Inflation) wird üblicherweise durch so genannte „Matched Model"-Methoden gewährleistet. Dabei wählen die statistischen Ämter eine Stichprobe von Gütern (Produkten) bei bestimmten Verkaufstellen (zur Bestimmung des Konsumentenpreisindex z.B. Einzelhändler, zur Bestimmung des Produzentenpreisindexes die Warenproduzenten selbst bzw. den Großhandel) und sammeln Preise für jedes einzelne Produkt in einer Anfangsperiode (Referenzzeitpunkt). In den darauf folgenden Perioden werden die Preise zu exakt denselben Produkten bei exakt denselben Verkaufstellen erneut notiert. Der Preisindex wird dann über das Matchen der jeweiligen Produkt-Modelle zu den entsprechenden Preisen der einzelnen Perioden im Bezug auf eine Referenzperiode ermittelt. Der Vorteil dieser Methode liegt darin, dass eine Vielzahl von preisbestimmenden Faktoren, die nicht beobachtbar oder direkt quantifizierbar sind, die aber auch keinen Inflationseffekt darstellen, dennoch konstant gehalten werden können.[12]

Qualitätsadjustierungen können alternativ auch mittels hedonischer Methoden durchgeführt werden. Dabei werden Preisregressionen geschätzt, bei denen die beobachteten Produktpreise als eine Funktion von Produktcharakteristika modelliert werden. Nicht explizit modellierte Effekte durch Variable für Produktcharakteristika, z.B. in Form von Dummys, können dann zur Bestimmung der Inflation herangezogen werden. Aus ihnen lassen sich anschließend hedonische Preisindizes konstruieren. Bespiele für hedonische Indizes, die direkt auf Preisregressionen zurückgehen, sind der Dummy-Variablen Ansatz, die Methode zur Ermittlung von

[12] Für eine detailiertere Betrachtung unterschiedlicher Matching Methoden siehe Triplett (2006).

Preisen für Charakteristika und die hedonische Qualitätsadjustierungs-Methode.[13] Grundsätzlich folgen alle Preisschätzungen unter Kontrolle einer Vielzahl unterschiedlicher Produktcharakteristika in Regressionsansätzen demselben Schema: Durch die Wahl des Regressionsansatzes können Effekte von Produktcharakteristika, die als Proxy-Variablen für Qualität verwendet werden, konstant gehalten und der Effekt der Preisbewegung (Inflation) isoliert betrachtet werden. Eine Trennung der beobachteten Preisentwicklung in eine reine Preis- und eine Qualitätskomponente kann somit gewährleistet werden.

2.6 Schlussfolgerung

Die Beschreibung der Unterschiede zwischen Volumen- und Preiskomponenten in diesem Kapitel hat versucht die herausragende Bedeutung von Preisen bei der Konstruktion von ökonomisch motivierten Indexzahlen deutlich zu machen. Preise sind von zentraler Bedeutung, insbesondere dann, wenn reale Outputindizes unter Berücksichtigung der ihnen innewohnenden Qualitätsveränderungen bestimmt und weit reichende wirtschaftspolitische Entscheidungen anhand der daraus resultierenden realen Daten getroffen werden sollen. So hatte der Boskin Report (1996) eine Überschätzung des U.S. Konsumentenpreisindizes (CPI) von 1,1 Prozentpunkten pro Jahr festgestellt, von denen 0,6 Prozentpunkte auf die Unterschätzung des Qualitätseffekts zurückgeführt werden konnten. Derartig deutliche Unterschiede in äußerst wichtigen Preisstatistiken machen die Dimension und die Notwendigkeit einer umfassenden Qualitätsadjustierung offensichtlich. Für den Öffentlichen Sektor ist die Thematik von verfügbaren Preisen zur Konstruktion geeigneter Outputindizes besonders gravierend, da dieser in vielen Fällen über keine eigenen Märkte verfügt, auf denen sich Preise bilden können. Die Bestimmung qualitätsadjustierter Volumenindizes nach den obigen Ausführungen, bei denen Qualitätskomponenten aus beobachtbaren Preisstatistiken extrahiert werden, ist somit nicht möglich.

[13] Für eine detailiertere Betrachtung unterschiedlicher Konstruktionen von hedonischen Preisindizes und alternativen Adjustierungsmethoden siehe Triplett (2006).

3. Probleme und Lösungsansätze der Outputmessung im Nicht-Markt Sektor

3.1 Einleitung und Problemstellung

Die Leistungen des Staates tragen in den meisten entwickelten Ländern einen nicht vernachlässigbaren Teil zum Bruttoinlandsprodukt (BIP) bei. Schon allein aus diesem Grund kommt der Messung der Produktion des öffentlichen Sektors eine bedeutende Rolle zu. Des Weiteren besteht ein öffentliches Interesse an einer effizienten Verwendung von Steuergeldern durch den Staat. Eine Möglichkeit, die Effizienz des öffentlichen Sektors zu messen, besteht in der Berechnung von Produktivitätskennziffern. Wie im vorangegangenen Kapitel erläutert, ergibt sich die Produktivität aus der Gegenüberstellung von realem Output zu realem Input. Während die Messung des realen Inputs im öffentlichen Sektor vergleichbaren Schwierigkeiten unterliegt wie im privatwirtschaftlichen Bereich, ergeben sich besondere Herausforderungen bei der Messung des staatlichen Outputs. Das Problem der Messung öffentlicher Produktion besteht darin, dass Transaktionen von Leistungen zwischen Staat und Konsumenten im Allgemeinen ohne Preise oder zu subventionierten Preisen stattfinden. Das bedeutet, dass selbst wenn Preise beobachtet werden können, diese nicht der Zahlungsbereitschaft und damit der Werteinschätzung der Konsumenten entsprechen.

Das vorangegangene Kapitel hat die zentrale Bedeutung der Preise für die Outputmessung hervorgehoben. Erstens ermöglicht die Beobachtung von Preisen, den nominalen Output eines Sektors oder einer Firma in eine Preis- und eine Volumenkomponente zu zerlegen. Darauf aufbauend können im Zeitablauf reale Veränderungen (reales Wachstum) des Outputs sowie die Veränderungen der Preise (Inflation) berechnet werden. Zweitens spiegeln Preise die marginale Zahlungsbereitschaft und damit die Werteinschätzung der Konsumenten gegenüber einem Produkt oder einer Leistung wider. Daher bilden Preise ökonomisch fundierte Gewichte bei der Aggregation heterogener Güter und Leistungen. Drittens resultieren Preisveränderungen neben der Inflation auch aus Veränderungen in der Qualität von Gütern und Dienstleistungen. Prinzipiell können durch geeignete Verfahren, wie den hedonischen Methoden, Preise in eine Inflations- und eine Qualitätskomponente zerlegt werden. Die Erfassung der Veränderung der Qualität von Gütern und Dienstleistungen ist eines der entscheidenden Prob-

leme bei der Messung des Outputs. Aufgrund der zentralen Bedeutung von ökonomisch relevanten Preisen, d.h. Preisen, die die marginale Zahlungsbereitschaft und/oder die marginalen Kosten widerspiegeln, stellt sich die Frage nach geeigneten alternativen Methoden für die Outputmessung, wenn Preisinformationen fehlen. Dieses Kapitel stellt diese Verfahren vor und diskutiert die jeweiligen Vor- und Nachteile.

Das Problem der Outputmessung im öffentlichen Sektor ist nicht neu. Es hat jedoch in der jüngsten Vergangenheit durch ein gestiegenes Interesse an Produktivitätskennzahlen des Staatssektors sowie durch verstärkte Anstrengungen und Bedürfnisse nach Harmonisierung statistischer Daten auf internationaler Ebene neue Aufmerksamkeit erlangt. Eine Reihe von internationalen Organisationen hat sich mit diesem Thema beschäftigt und Empfehlungen formuliert. Zunächst ist hier das *System of National Accounts* (SNA) in der Version 1993 von den Vereinten Nationen zu nennen, welches das Problem der Messung des öffentlichen Outputs explizit anerkennt. Das SNA empfiehlt, dass die Messung des öffentlichen Outputs im Wesentlichen und soweit möglich den Konzepten und Prinzipien des privaten Sektors folgen soll. Das SNA unterscheidet explizit *individuelle* und *kollektive* Güter und unterstreicht die besonderen Schwierigkeiten der Messung des Outputs bei *kollektiven* Gütern. Das *European System of Accounts* (ESA) von 1995 greift diese Empfehlungen auf und passt sie den besonderen Gegebenheiten und Datenbedürfnissen der Europäischen Union an. Besondere Bedeutung kommt hierbei dem *Eurostat Handbook on Price and Volume Measures in National Accounts* (Eurostat, 2001) zu. In diesem Handbuch werden verschiedenen Verfahren der Outputmessung diskutiert sowie eine Klassifizierung der Ansätze in A/B/C-Methoden vorgenommen. Nach dieser Klassifikation bilden A-Methoden die geeignetsten und empfohlenen Ansätze. B-Methoden sind zulässig, wenn eine Umsetzung von A-Methoden nicht möglich ist. C-Methoden werden grundsätzlich als unzulässig eingestuft. Die Empfehlungen des Handbuches wurden in einer Entscheidung der Europäischen Kommission des Jahres 2002 zum Ausdruck gebracht. Darin wurden die Mitgliedsstaaten angehalten, ihre Methoden nach dem A/B/C-Schema zu klassifizieren sowie die Empfehlungen bis zum Jahr 2006 umzusetzen. In Deutschland wurden diese Empfehlungen vom Statistischen Bundesamt in Teilen in der Revision der Volkswirtschaftlichen Gesamtrechnung des Jahres 2005 umgesetzt (vgl. Braakmann et al., 2005).

Die größte Aufmerksamkeit im Bereich der öffentlichen Outputmessung ist dem so genannten *Atkinson Review* (Atkinson, 2005) zu teil geworden. Dieser Report bezieht sich auf die Situa-

tion der Outputmessung im Vereinigten Königreich und enthält die bei weitem konkretesten Empfehlungen der Outputmessung für die verschiedenen Bereiche des öffentlichen Sektors. Aus diesem Grund hat der Report auch auf internationaler Ebene große Bedeutung erlangt. Des Weiteren ist in diesem Kontext das *OECD Productivity Manual* (OECD, 2001) zu nennen, welches allgemeine Richtlinien zur Messung der Produktivität aufstellt und einen starken Bezug zur ökonomischen Theorie herstellt. Die OECD ist zudem aktuell damit beschäftigt, ein Handbuch zur Messung des Outputs im Bildungs- und Gesundheitsbereichs zu erstellen (vgl. OECD, i.E.).

Die nachfolgenden Ausführungen sind wie folgt gegliedert: Zunächst erfolgt eine Klassifizierung des öffentlichen Outputs in *individuelle* und *kollektive* Leistungen (vgl. Abschnitt 3.2). Diese Unterscheidung ist zum Verständnis der Anwendbarkeit der verschiedenen Methoden der Outputmessung zweckmäßig. In Abschnitt 3.3 werden diese zur Deflationierungsmethode alternativer Verfahren zur Berechnung des realen Outputs diskutiert. In diesem Zusammenhang sind insbesondere die Methoden der *direkten Outputmessung* zu nennen. Des Weiteren stellt sich die Frage nach einer geeigneten Aggregation oder Gewichtung von heterogenen Leistungen, wenn Werteinschätzungen der Konsumenten gegenüber verschiedenen Leistungen nicht direkt über die Preise beobachtbar sind (vgl. Abschnitt 3.4). Eines der wichtigsten und schwierigsten Probleme im Zusammenhang mit der Outputmessung bildet die Erfassung von Qualitätsveränderungen öffentlicher Güter und Dienstleistungen. Verfahren zur Qualitätsadjustierung sind Gegenstand des Abschnitts 3.5. Abschnitt 3.6 fasst die zentralen Ergebnisse dieses Kapitels zusammen.

3.2 Kollektive vs. individuelle Güter

Individuelle Güter und Dienstleistungen sind Leistungen, die von einzelnen Haushalten nachgefragt und konsumiert werden. Haushalte oder Individuen können andere Haushalte oder Individuen prinzipiell von der Nutzung der Leistung ausschließen (Exkludierbarkeit). Hierzu zählen beispielsweise Bildungs-, Gesundheits- und Sozialversicherungsleistungen. Entscheidend ist, dass im Fall der individuellen Güter eine konkrete Transaktion beobachtbar ist und somit der Output prinzipiell definierbar und auch statistisch erfassbar ist. Insofern sind individuelle Güter des öffentlichen Sektors traditionellen Gütern, die auf einem Markt angeboten werden, vergleichbar. Der Hauptunterschied besteht darin, dass ein Preis bei den öffentlichen Gütern nicht beobachtet werden kann.

Neben *individuellen* Leistungen werden vom Staat auch so genannte *kollektive* Dienstleistungen bereitgestellt. *Kollektive* Dienstleistungen stehen der Gesellschaft als ganzes zur Verfügung. Ein einzelner Haushalt ist nicht in der Lage andere Haushalte von der Nutzung solcher Dienstleistungen auszuschließen (keine Exkludierbarkeit). Als Beispiele für kollektive Güter sind die Verteidigung eines Landes sowie präventive Maßnahmen etwa in der Gesundheitsvorsorge zu nennen. Bei kollektiven Gütern ist eine konkrete Transaktion zwischen Anbietern und Nachfragern prinzipiell nicht beobachtbar und es ist somit problematisch den Output zu definieren und zu messen.

Erschwerend kommt hinzu, dass es auch eine Reihe von Fällen gibt, in denen Anbieter von öffentlichen Leistungen sowohl *kollektive* als auch *individuelle* Leistungen erbringen. Als ein Beispiel seien hier die Leistungen der Feuerwehr zu nennen. Ein Feuerwehreinsatz zur Brandbekämpfung in einer Wohnung kann als *individuelle* Leistung aufgefasst werden. Im Gegensatz dazu sind Maßnahmen zur Brandprävention als kollektive Leistungen zu klassifizieren.

Die Unterscheidung in *individuelle* und *kollektive* Leistungen hat Konsequenzen für die Möglichkeiten der Outputvolumenmessung. Da bei *kollektiven* Dienstleistungen ein konkreter Output meist nicht definierbar ist, muss auf traditionelle Inputverfahren zur Messung des Outputs zurückgegriffen werden. Im Gegensatz dazu können bei *individuellen* Gütern so genannte *direkte* Outputmessverfahren verwendet werden. Diese Unterscheidung spiegelt sich auch in der unterschiedlichen Klassifizierung von Verfahren in A/B/C-Methoden durch Eurostat wider. Während Inputverfahren generell als C-Methoden klassifiziert werden und somit zu vermeiden sind, werden sie im Falle von kollektiven Gütern als B-Methoden und damit als ggf. akzeptabel eingestuft. Die verschiedenen Verfahren der Outputmessung sind Gegenstand des folgenden Abschnitts.

3.3 Verfahren der Outputvolumenmessung

Wie bereits im vergangenen Kapitel erläutert, ist es zweckmäßig, die Veränderung des nominalen Outputs in eine Volumen- und eine Preis- oder Inflationskomponente zu zerlegen. Diese Zerlegung ist beispielsweise entscheidend für die Berechnung von Produktivitätskennzahlen, welche ausschließlich auf realen Größen basieren. Bei Gütern und Dienstleistungen, die zu Marktpreisen gehandelt werden, wird diese Zerlegung vorzugsweise durch die Deflationierungsmethoden mittels geeigneter Preisindizes erreicht (vgl. Abschnitt 1.4). Da per Definition solche Preisindizes bei öffentlichen Gütern und Dienstleitungen nicht existieren, müssen alternative Möglichkeiten der Outputvolumenmessung in öffentlichen Sektoren gefunden werden. Grundsätzlich bestehen hierfür zwei Möglichkeiten. Das erste Verfahren basiert auf einer Approximation des realen Outputs durch die deflationierten Inputs. Das zweite Verfahren versucht, das Outputvolumen *direkt* über geeignete Indikatoren zu messen. Im Folgenden werden die Verfahren kurz skizziert und diskutiert. Abbildung 2.1 stellt den Zusammenhang dieser Verfahren schematisch dar.

3.3.1 Inputverfahren

Die traditionelle Methode ist das so genannte *Inputverfahren*. Dieses Verfahren wird z.B. in den USA weiterhin zur Outputmessung im öffentlichen Sektor eingesetzt. Bei dieser Methode werden einige ausgewählte oder alle Inputs eines Sektors (Kapital, Arbeit und Vorleistungen) mit entsprechenden Preisindizes deflationiert und mit dem realen Output gleichgesetzt. Der Vorteil dieses Verfahrens besteht in der einfachen Implementierbarkeit, da generell die Informationen zu den Inputfaktoren vorliegen. Die Nachteile sind jedoch offensichtlich: Da per definitionem Veränderungen des Volumens der eingesetzten Inputfaktoren den Veränderungen des Volumens des Outputs entsprechen, können Produktivitätsveränderungen, die z. B. durch den effizienteren Einsatz der Inputfaktoren oder technischen Fortschritt entstehen können, nicht erfasst werden. Um diesem Problem zu begegnen, haben einige Länder exogen eine Produktivitätsverbesserung angenommen. So hat beispielsweise das Statistische Bundesamt in Deutschland bis zur Revision des Jahres 2005 eine Produktivitätssteigerungsrate von 0.5% für die Sektoren des öffentlichen Bereiches angenommen. Exogene Annahmen über Produktivitätsverbesserungen sind jedoch willkürlich, wenn sie jeder empirischen Grundlage entbehren. Die Interpretation des Produktivitätsmaßes hängt zudem davon ab, ob nur ein Input oder alle Inputfaktoren zur Berechnung des Outputs herangezogen werden. Basiert das Outputmaß le-

diglich auf dem Arbeitsinput, so wird angenommen, dass die *Arbeitsproduktivität* konstant ist. Werden hingegen alle Inputs zur Berechnung des Outputs verwendet, wird angenommen, dass die *Totale Faktorproduktivität* konstant bleibt. In diesem zweiten Fall kann es zu einer Steigerung der *Arbeitsproduktivität* aufgrund steigender Kapitalausstattung jedes Arbeitnehmers kommen[14]. Des Weiteren besteht ein Problem in der Erfassung von Qualitätsverbesserungen. Generell können Qualitätsverbesserungen indirekt über Qualitätssteigerungen der Inputfaktoren erfasst werden. Dies beinhaltet jedoch die implizite Annahme, dass Qualitätssteigerungen der Inputfaktoren zu einer gleichwertigen Steigerung der Qualität des Outputs führen. Diese Annahme kann jedoch nicht verifiziert werden, ohne die Qualität des Outputs selber zu messen.

Aufgrund der konzeptionellen Schwierigkeiten hat die Europäische Kommission in ihrer Entscheidung vom Dezember 2002 die Inputmethode als C-Methode klassifiziert und deren Verwendung ab dem Jahre 2006 für nicht mehr zulässig erklärt. Stattdessen soll der Output einer öffentlichen Einrichtung direkt gemessen werden. Ausnahme bilden öffentliche Sektoren, die vorwiegend durch kollektive Leistungen charakterisiert sind. Der Grund besteht darin, dass der Output bei kollektiven Gütern besonders schwer zu messen ist. Beispielsweise ist es problematisch, den Output der Landesverteidigung oder der öffentlichen Verwaltung zu definieren. Im Gegensatz dazu ist bei individuellen Leistungen eine Transaktion zwischen dem Staat als Anbieter und dem Bürger als Konsumenten beobachtbar und daher der Output prinzipiell definierbar.

3.3.2 Direkte Outputvolumenmessung

Bei den Verfahren der direkten Outputvolumenmessung wird versucht, die Leistungen des öffentlichen Sektors direkt über geeignete Indikatoren zu erfassen. Wie bereits im Abschnitt 1.4 erläutert, sollte sich ein geeignetes Outputvolumenmaß aus einer Mengen- und einer Qualitätskomponente zusammensetzen. Wie im Verlaufe dieses Abschnitts deutlich werden wird, ist es aus diesem Grund zweckdienlich, eine Unterscheidung in Aktivitäten, Output und Outcome zu vollziehen (Eurostat, 2001).

[14] Vergleiche hierzu auch die Ausführungen zur Einzelfaktor- und Multifaktorproduktivität des Abschnitts 2.3.

Aktivitäten

Aktivitäten sind Tätigkeiten oder Prozesse, die ein Produzent zur Erstellung einer Leistung mit den gegebenen Inputfaktoren ausführt. Insofern stehen sie in engerem Zusammenhang zum Output als die Inputfaktoren. Beispiele für Aktivitäten im öffentlichen Sektor sind die Anzahl an Operationen in einem Krankenhaus und die Anzahl an Schulstunden und/oder Vorlesungs- und Seminarzeiten im Bildungssystem. Generell liegen Daten zu einer Vielzahl von Aktivitäten des öffentlichen Sektors vor. Aus diesem Grund liegt es nahe, die Leistungen eines öffentlichen Sektors durch die Anzahl der Aktivitäten zu messen. Dieses Vorgehen birgt jedoch mindestens zwei Gefahren.

Zum einen besteht das Problem, dass ein Sektor durch eine Vielzahl von Aktivitäten gekennzeichnet ist. Um ein unverzerrtes Bild des Outputs eines Sektors zu geben, müssten prinzipiell alle Aktivitäten erfasst werden. Ist dies aus Gründen der Datenverfügbarkeit nicht möglich, so hängt die Güte der Outputmessung entscheidend davon ab, ob die erfassten Aktivitäten als stellvertretend für die Aktivitäten des gesamten Sektors angesehen werden können. Beispielsweise könnte die Anzahl der Schulstunden in Gymnasien im Zeitablauf stark gestiegen sein, wären die Schulstunden in Realschulen gesunken ist. Liegen nur Daten zu Schulstunden im Gymnasialbereich vor und werden diese Stunden als Indikator des Outputs des Bildungsbereichs verwendet, so ergibt sich ein stark verzerrtes Bild der Gesamtleistung des Bildungsbereiches.

Ein zweites Problem besteht darin, dass auf Aktivitäten basierende Outputmaße Qualitätsveränderungen nicht adäquat erfassen können. Die Qualität einer Leistung definiert sich über einen Satz von Merkmalen oder Charakteristika. Qualitätsveränderungen treten somit dann auf, wenn neue Produkte auf einem Markt eingeführt werden oder sich die Charakteristika der bestehenden Produkte im Zeitablauf verändern. Um das Problem der Qualitätsveränderungen bei der Outputmessung durch Aktivitäten zu illustrieren, sei auf ein Beispiel verwiesen. Es sei angenommen, dass eine neue medizinische Behandlungsmethode eine traditionelle Behandlungsmethode im Zeitablauf ersetzt. Es sei weiter angenommen, dass die neue Behandlungsmethode weniger Operationen benötigt als die traditionelle, um zum gleichen Behandlungserfolg zu gelangen. Würde die Zahl der Operationen als Aktivitätsindikator verwendet, ergäbe sich eine Reduktion des Outputs im Zeitablauf. Dies erscheint in dem vorliegenden Fall jedoch nicht plausibel.

Im Bereich der Marktproduktion führen Qualitätsveränderungen generell zu Preisveränderungen, da veränderte Produktcharakteristika Veränderungen der Kosten und Zahlungsbereitschaften implizieren. Über hedonische Methoden können den einzelnen Produktmerkmalen Marktbewertungen zugeordnet werden. Durch die Aggregation der einzelnen Produktmerkmale mit ihren Marktbewertungen lässt sich dann ein sinnvolles Qualitätsmaß berechnen. Im öffentlichen Sektor stehen solche Marktinformationen nicht zur Verfügung. Daher müssen in diesem Bereich alternative Methoden zur Bewertung von Produktcharakteristika gefunden werden. Hierzu eignet sich das Konzept des Outcomes.

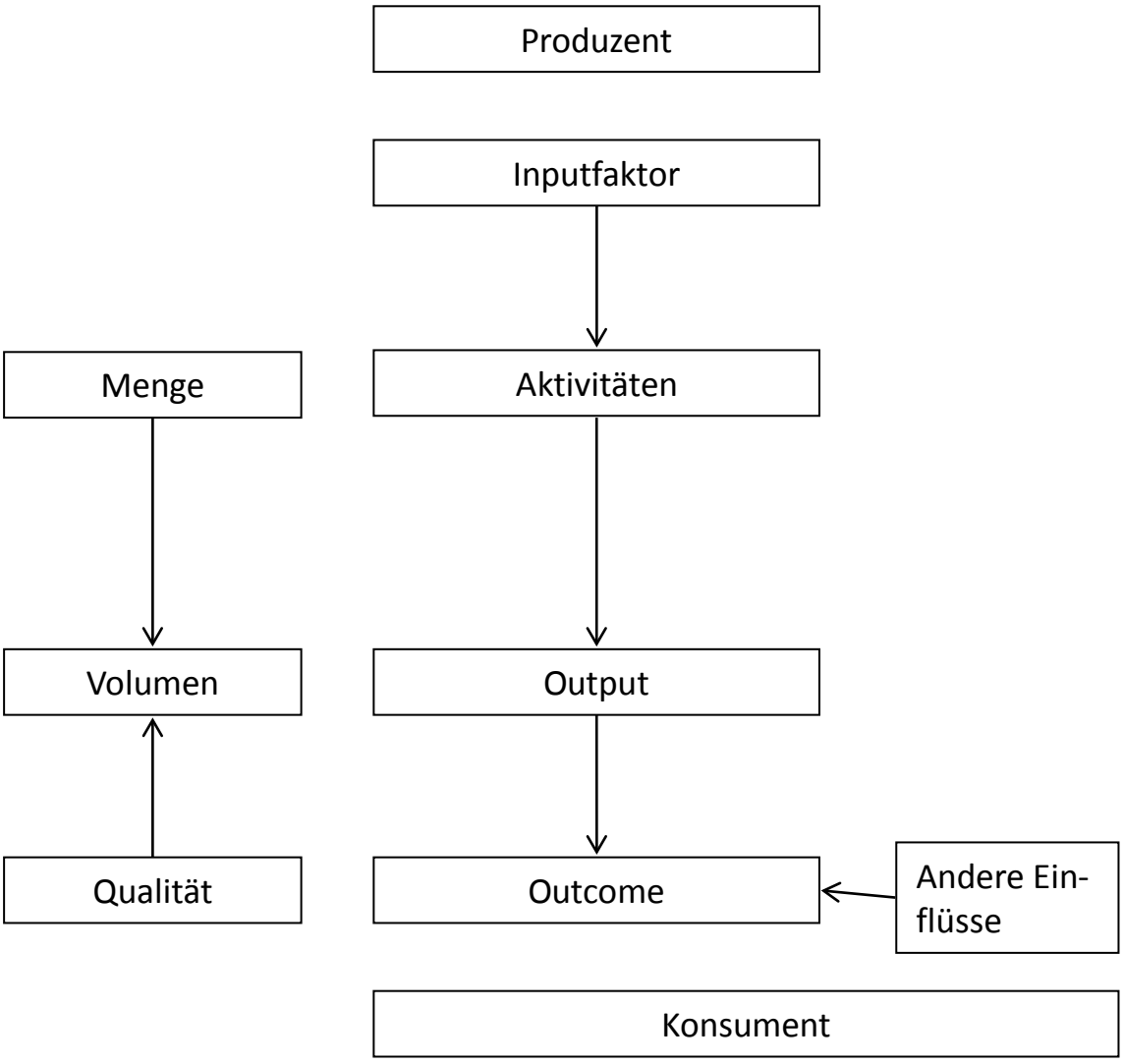
Outcome

Der Outcome misst einen Zustand, der das Nutzenniveau eines Konsumenten beeinflusst. Beispiele für Outcome Maße sind der Bildungsstand eines Individuums gemessen beispielsweise durch die Ergebnisse von Leistungstests oder der Gesundheitszustand gemessen durch die Lebenserwartung. Ein Konsument kann aus dem Bildungs- oder Gesundheitszustand direkt einen Nutzen ziehen, oder indirekt durch den Effekt auf zukünftiges Einkommen und Konsum. Der Outcome steht daher im Gegensatz zu Aktivitäten im direkten Zusammenhang zum Nutzen der Konsumenten. Allerdings wird der Outcome neben den Aktivitäten einer öffentlichen Einrichtung auch durch eine Reihe von anderen Faktoren beeinflusst werden. Beispielsweise hängt der Bildungsstand einer Person neben den konsumierten Leistungen des Bildungssektors auch von den Erbanlagen, dem sozioökonomischen Hintergrund sowie von eigenen Anstrengungen ab. Der Gesundheitszustand wird neben den konsumierten Leistungen des Gesundheitssystems auch u.a. von der Umwelt und vom eigenen Lebensstil beeinflusst. Aus diesem Grund eignen sich Outcome-Indikatoren in den meisten Fällen nicht direkt zur Outputmessung.

Outcome Maße geben jedoch wichtige Hinweise über die Bewertung verschiedener Aktivitäten durch die Konsumenten. Sie ermöglichen es somit, Aussagen über die Qualität zu treffen. Einer Aktivität, die in stärkerem Maße zur Verbesserung eines Outcomes beiträgt als eine andere Aktivität, wird von Konsumenten eine höhere Qualität beigemessen. Ebenso können verschiedene Charakteristika einer Aktivität bewertet werden, indem ihre Beiträge auf die Veränderung des Outcomes hin untersucht werden. Beispielsweise könnte analysiert werden, wie die Klassengröße oder die Ausstattung mit Computern die Ergebnisse von Leistungstests im Schulsektor beeinflussen. Ähnlich wie Preise im Marktsektor bilden Outcome Maße daher

ein aggregiertes Qualitätsmaß. Sie sind entscheidender Baustein für die Qualitätsadjustierung, wenn Preissignale fehlen (vgl. Abschnitt 3.5).

Abb.: 3.1: Schematische Darstellung der Outputvolumenmessung



Output

Der Output steht zwischen Aktivitäten und Outcome. Idealerweise misst der Output den direkt zurechenbaren Einfluss der Aktivitäten eines Produzenten auf den Outcome. Der Output des Bildungssektors kann beispielsweise als der Transfer von Wissen und Fähigkeiten definiert werden. Es soll also der Beitrag zur Verbesserung des Bildungsstandes (Outcome) gemessen werden, welcher direkt auf die Aktivitäten des Bildungssektors (z.B. Schulstunden) zurückgeführt werden kann. Outputindikatoren lassen sich somit aus der Kombination von Aktivitätsindikatoren und Outcome-Indikatoren erstellen, wobei Aktivitätsindikatoren die

reine Quantitätskomponente des Outputs darstellen und Outcome-Indikatoren die Qualitätskomponente erfassen. Aus diesem Grund definiert die OECD Outputindikatoren auch als qualitätsadjustierte Aktivitätsindikatoren (OECD, i.E.).

Die Unterscheidung in Input/Aktivitäten/Output und Outcome spiegelt sich auch in der Klassifikation in A/B/C-Methoden durch Eurostat wieder. Zur Outputmessung von *individuellen Leistungen* gelten nur Outputindikatoren (qualitätsadjustierte Aktivitätsindikatoren) als A-Methoden. Reine Aktivitätsindikatoren, die keine Qualitätsveränderungen berücksichtigen, können maximal als B-Methoden klassifiziert werden. Inputverfahren sind grundsätzlich C-Methoden.

3.4 Aggregation heterogener Gruppen

Der Output eines Sektors besteht aus einer Vielzahl heterogener Güter und Dienstleistungen. Beispielsweise umfasst der Output des Gesundheitswesens so unterschiedliche Dienstleistungen wie Herztransplantationen und Zahnfüllungen. Aus ökonomischer Sicht ist es nicht sinnvoll diese verschiedenen Leistungen einfach zu summieren, um auf ein aggregiertes Volumenmaß für den sektoralen Output zu gelangen, da diese Leistungen sowohl sehr unterschiedliche Kosten der Anbieter als auch unterschiedliche Werteinschätzungen der Konsumenten widerspiegeln. Es stellt sich also die Frage nach einer sinnvollen Gewichtung heterogener Leistungen.

Wie bereits im vorangegangenen Kapitel diskutiert, ist diese Frage im Bereich der Marktleistungen eindeutig beantwortet. Relative Preise sollten als Gewichte dienen, da diese die marginalen Werteinschätzungen (Grenznutzen) der Konsumenten widerspiegeln. Ein Produkt erhält somit ein höheres Gewicht, wenn es bei einem Konsumenten mit einem höheren marginalen Nutzen verbunden ist. Bei öffentlichen Leistungen besteht nun wiederum das Problem, dass Preise nicht beobachtbar sind und somit der marginale Nutzen nicht ableitbar ist. Neben der Unbeobachtbarkeit der Preise kommt hinzu, dass es aus theoretischen Gründen kaum Veranlassung gibt zu postulieren, dass eine öffentliche Leistung bis zu dem Punkt bereitgestellt wird, an dem ein Gleichgewicht zwischen marginalem Nutzen den marginalen Kosten einstellt. Abbildung 3.2 verdeutlicht diesen Zusammenhang. Unter der Annahme der vollständigen Konkurrenz stellt sich eine Gleichgewichtsmenge q^c ein. An diesem Punkt entspricht der Preis p^c gleichzeitig der marginalen Werteinschätzung und den marginalen Kosten. Das be-

deutet, dass selbst wenn der Preis unbeobachtbar wäre, die (marginalen) Kosten des Produzenten als Maß für die marginale Werteinschätzung herangezogen werden können. Grundsätzlich besteht jedoch kein Grund anzunehmen, dass die bereitgestellte Menge eines öffentlichen Gutes dieser Gleichgewichtsmenge entspricht. In der Abbildung sei angenommen, dass das öffentliche Gut in der Menge q^s angeboten wird. In diesem Punkt fallen marginale Kosten und Werteinschätzungen auseinander. In unserem Beispiel liegen die marginalen Kosten unter dem marginalen Nutzen. Daher liefern die Kosten des Anbieters ein verzerrtes Bild über die Werteinschätzungen der Konsumenten. Dies hat Auswirkungen auf die Volumenmessung. Beispielsweise würde eine Ausweitung der Menge des Gutes zu einer Erhöhung des Gewichts dieses Gutes bei Kosten-Gewichtung und zu einer Verringerung des Gewichts bei Nutzengewichtung führen.

Hypothetische Preise

Um mit dem Problem der Gewichtung umzugehen, gibt es verschiedene Lösungsmöglichkeiten. Zunächst können *hypothetische Preise* aus alternativen oder parallelen Märkten als approximative Größe abgeleitet werden. Ein Beispiel für einen alternativen Markt wäre der Schienenverkehr für den Straßenverkehr. Zwischen Straßen- und Schienenverkehr besteht eine enge Substitutionsbeziehung aus Sicht der Konsumenten. So könnten für die Preise für den Personen- oder Frachtverkehr auf der Strasse äquivalente Preise des Schienenverkehrs zu Rate gezogen werden. Als Beispiel für die Existenz eines Parallelmarktes sind der Gesundheitsbereich und hier insbesondere der Pflegebereich zu nennen. Hier können die Preise privater Anbieter für Pflegedienstleistungen verwendet werden, um zu einer Bewertung öffentlicher Pflegedienstleistungen zu gelangen. Das Problem paralleler Märkte für öffentliche Güter besteht jedoch häufig darin, dass sie in hohem Maße reguliert sind. Daher entsprechen die Preise in diesen Märkten nur selten der wahren Zahlungsbereitschaft und damit der Werteinschätzung der Konsumenten.

Outcome basierte Gewichtung

Eine weitere Möglichkeit besteht darin, die Werteinschätzungen durch die Verwendung von Outcome-Indikatoren abzuleiten. Wie im vorangegangenen Abschnitt beschrieben, messen Outcome-Indikatoren den Erfolg verschiedener Aktivitäten auf nutzenrelevante Größen der Konsumenten. Das Problem besteht darin, dass dieser Erfolg auch durch Faktoren beeinflusst

wird, die außerhalb des Beeinflussungsbereichs der öffentlichen Einrichtungen liegen. Sind alle Einflussfaktoren des Outcomes bekannt, so lässt sich theoretisch der partielle Einfluss verschiedener öffentlicher Aktivitäten auf den Outcome unverzerrt schätzen. Ein höherer Beeinflussungsgrad einer Aktivität gegenüber einer zweiten entspricht nach dieser Logik einer höheren Wertschätzung des Konsumenten, da diese Aktivität den Nutzen in stärkerem Maße beeinflusst. Daher können partielle Beeinflussungsgrade als theoretisch sinnvolle Gewichte dienen. Für ein Beispiel sei wiederum auf den Bildungsbereich verwiesen. Das erwartete zukünftige Einkommen kann als *ein* sinnvolles Outcome Maß des Bildungsbereiches angesehen werden. Das zukünftige Einkommen wird *unter anderem* von der maximal erreichten Bildungsstufe eines Individuums bestimmt (z.B. Primär-, Sekundär- oder Tertiärbildung). Es lassen sich somit so genannte Lohnprämien schätzen, d.h. der zusätzliche erwartete Lohn einer Bildungsstufe gegenüber einer Referenzbildungsstufe. Diese Lohnprämien können wiederum als Gewichte verwendet werden, um den Output verschiedener Bildungsstufen (z.B. gemessen durch die Schüler- und Studentenzahlen) zu einem Gesamtmaß des Bildungssektors zu aggregieren.

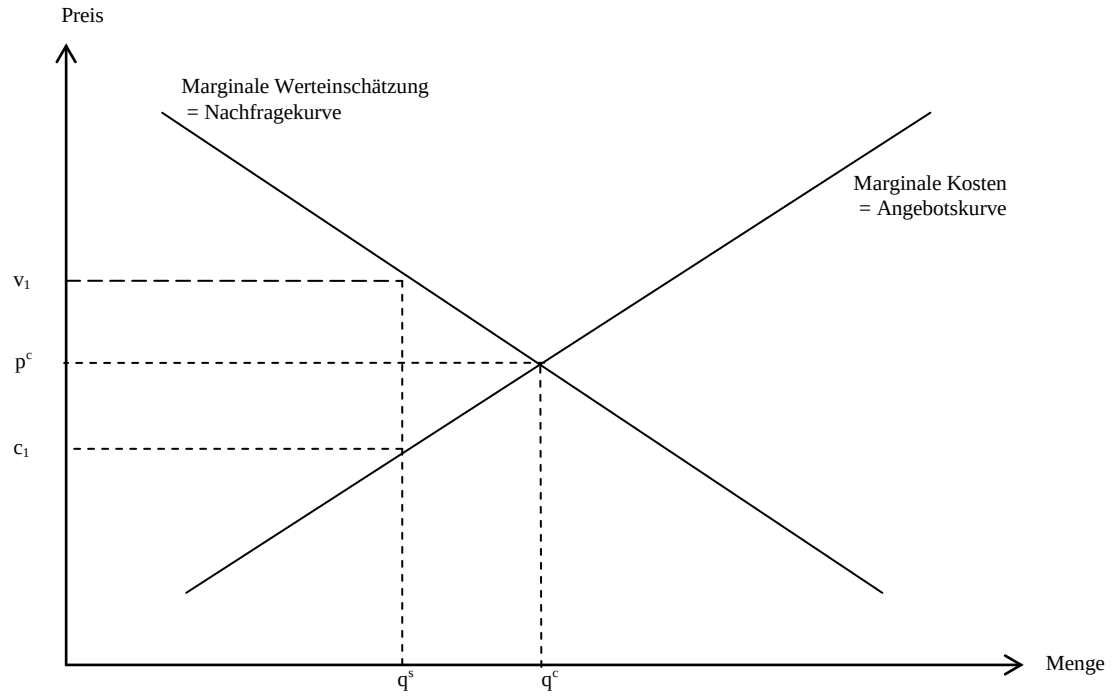
Kostengewichte

Als Alternative zu hypothetischen Preisen oder Outcome basierten Gewichten können auch Kosten verwendet werden, um die Zahlungsbereitschaft der Konsumenten zu approximieren. Wie bereits erläutert besteht generell kein Anlass anzunehmen, dass die Kosten der Bereitstellung eines öffentlichen Gutes den Werteinschätzungen der Konsumenten entsprechen (vgl. Abbildung 3.2). Es lässt sich jedoch argumentieren, dass Konsumenten in einer Demokratie durch den Wahlprozess Einfluss auf die bereitgestellte Menge nehmen können. In diesem Fall kann angenommen werden, dass die bereitgestellte Menge dem sozialen Optimum zumindest langfristig entspricht. Das soziale Optimum wäre bei der Gleichgewichtsmenge q^c in Abbildung 3.2 erreicht, da in diesem Punkt die Summe aus Konsumenten- und Produzentenrente maximiert wird. Folgt man dieser Argumentation, so ergibt sich daraus, dass die marginalen Kosten der marginalen Zahlungsbereitschaft entsprechen und somit Kosten als Gewichte verwendet werden können (OECD,i.E.).

Ein Problem bei der Verwendung von Kostengewichten tritt bei der Einführung neuer kosteneffizienterer Verfahren auf, die parallel zu traditionellen Methoden existieren. Unter der Annahme, dass die neue und die traditionelle Methode aus Sicht des Konsumenten einen ver-

gleichbaren Output liefern, führt die Verwendung der Kostengewichtung zu einer Unterschätzung des aggregierten Outputs (Diewert, 2008). Der Grund besteht darin, dass der Output der kosteneffizienteren Methode bei Kostengewichtung ein geringeres Gewicht erhält als der Output der traditionellen Methode, obwohl der Output annahmegemäß vom Konsumenten als gleichwertig eingeschätzt wird. Als Beispiel kann hier eine neue kosteneffizientere Therapie zur Behandlung einer Krankheit angeführt werden. Dieses Problem kann dadurch gelöst werden, dass Verfahren, die aus Konsumentensicht äquivalenten Output generieren, als homogene Leistungen aufgefasst werden. Eine Klassifikation in homogene und heterogene Güter sollte demnach ausschließlich auf Grundlage der Substitutionsbeziehung verschiedener Leistungen aus Sicht der Konsumenten erfolgen und nicht auf den Kosten basieren. Ein „korrektes" Outputmaß bei homogenen Produkten ergibt durch die ungewichtete Summierung der Outputs.

Abb.: 3.2: Schematische Darstellung des Gleichgewichts zwischen marginalem Nutzen und Kosten



Ein weiteres Problem der Verwendung von Kostengewichten besteht in der Wahl des richtigen Kostenmaßes. Wie Abbildung 3.2 verdeutlicht sollten marginale Kosten, d.h. die Kosten, die durch die Produktion einer zusätzlichen Einheit des Outputs anfallen, verwendet werden. Häufig stehen diese Informationen jedoch nicht zur Verfügung, so dass auf die durchschnittlichen Kosten einer Einheit (*unit costs*) zurückgegriffen wird. Marginale und Durchschnittskos-

ten entsprechen sich jedoch nur in einem Spezialfall: Wenn für die Produktionstechnologie des Anbieters konstante Skalenerträge angenommen werden, d.h. dass eine Vervielfachung jedes Inputfaktors um den Faktor x eine Erhöhung des Gesamtoutputs um den Faktor x gegenübersteht.[15]

3.5 Qualitätsmessung

Ein weiteres Problem der Outputmessung besteht darin, dass die Qualität des Outputs im Zeitablauf nicht konstant bleibt. Neue Produkte treten auf den Markt oder die Qualitätsattribute bestehender Leistungen ändern sich. Für unsere Zwecke relevant ist, dass sich Qualitätsveränderungen in einer Veränderung der Zahlungsbereitschaft der Konsumenten widerspiegeln. Bei Gütern und Dienstleistungen, die auf Märkten gehandelt werden, lassen sich Qualitätsveränderungen somit durch Preisveränderung beobachten. Die Hauptschwierigkeit besteht dann darin, die Preisveränderung einer Leistung in eine Qualitätskomponente und eine Inflationskomponente zu zerlegen. Dies ist kein triviales Problem und ist Gegenstand intensiver Forschung. Hierzu sind eine Reihe von Methoden entwickelt worden, die teilweise im vorangegangenen Kapital diskutiert wurden. In Sektoren, in denen solche Preisinformationen, die Rückschlüsse auf die Zahlungsbereitschaft zulassen, nicht vorliegen, also insbesondere im öffentlichen Sektor, müssen andere Verfahren verwendet werden. Wie im Folgenden genauer dargelegt, spielen in diesem Zusammenhang die oben beschriebenen Outcome-Maße eine besondere Rolle.

Eine Möglichkeit der Qualitätserfassung besteht darin, die verschiedenen Leistungen eines Sektors so tief wie möglich zu differenzieren. Eine Qualitätsveränderung würde sich in diesem Fall über eine Veränderung der Proportionen der einzelnen homogenen Leistungsgruppen erfassen lassen (so genannter *Qualitätsstruktureffekt*). Die Zweckmäßigkeit dieses Ansatzes hängt jedoch in entscheidendem Maße von der Wahl der Gewichte zur Aggregation in einem Gesamtoutputindex ab. Nur wenn Leistungen von höherer Qualität auch ein höheres Gewicht im Gesamtindex erhalten, ergibt sich durch eine Verschiebung hin zu Outputs besserer Qualität auch der gewünschte Effekt der Steigerung des Gesamtoutputs eines Sektors. Dies lässt sich z.B. durch die Verwendung von Gewichten erreichen, die den Einfluss verschiedener Leistungen auf ein geeignetes Outcome-Maß widerspiegeln. Beispielsweise könnte die Quali-

[15] In der Abbildung 3.2 haben wir keine konstanten Skalenerträge angenommen. Dies kommt in der Steigung der marginalen Kostenkurve zum Ausdruck. Konstante Skalenerträge würden eine konstante (waagerechte) marginale Kostenkurve implizieren.

tät unterschiedlicher Schulen durch Leistungstests gemessen werden. Eine Verschiebung der Schülerzahlen hin zu Schulen, die durchschnittlich besser bei Leistungstests abschneiden, würde so zu einer Erhöhung des qualitätsadjustierten Gesamtoutputs des Schulsystems führen. Im Gegensatz dazu können Kostengewichte zu einem verzerrten Bild führen. Eine kostengünstigere Therapie könnte beispielsweise zu einem gleichen oder verbesserten Behandlungsergebnis (Outcome) führen als eine kostenintensivere. Eine Verschiebung hin zu dieser kostengünstigeren Therapie würde sich allerdings in einer Reduktion des Gesamtoutputs niederschlagen.

Eine weitere Möglichkeit besteht darin, die Qualität einer Leistung direkt in der Definition des Quantitätsmaßes zu erfassen. In diesem Fall wird versucht, anstatt durch Aktivitätsindikatoren den Output direkt über Outcome-Indikatoren zu messen. Beispielsweise könnte der Output der Universitäten, anstatt durch die Anzahl der Studenten (Aktivitätsindikator) durch die Anzahl der Absolventen (Outcome-Indikator) ausgedrückt werden. Erhöht sich die Anzahl der Absolventen bei gleich bleibender Zahl der Studenten, so kann dies als Qualitätsverbesserung des Universitätssektors aufgefasst werden. Wie bereits in Abschnitt 3.3 angesprochen, besteht jedoch bei Outcome-Indikatoren das Problem der Zurechenbarkeit. Outcome-Indikatoren werden in der Regel auch von Faktoren beeinflusst, die nicht direkt im Zusammenhang mit der Dienstleistung der öffentlichen Einheit stehen. Beispielsweise wird die Absolventenrate unter anderem auch von individuellen Charakteristika der Studenten beeinflusst. Es bleibt also festzuhalten, dass sich die Methode der Qualitätsberücksichtigung durch Erfolgsindikatoren nur dann eignet, wenn der Outcome den Leistungen eines öffentlichen Sektors eindeutig zugerechnet werden kann.

Zeichnet sich eine Leistung durch eine Vielzahl von Qualitätsattributen aus, so bietet sich eine weitere Möglichkeit der Qualitätsadjustierung an. Hierbei wird das Volumen des Outputs weiterhin über Aktivitätsindikatoren gemessen, jedoch wird versucht, den Zusammenhang zwischen den Charakteristika einer Aktivität zum Outcome Maß zu quantifizieren und durch eine Qualitätsadjustierung zu berücksichtigen. Auf Grundlage der Zusammenhänge zwischen den Leistungsmerkmalen und dem Erfolgsindikatoren werden die Aktivitätsindikatoren dann herauf- oder herunterskaliert. Ähnlich wie bei hedonischen Methoden müssen zunächst Qualitätsattribute identifiziert werden und genaue Analysen durchgeführt werden, welche Faktoren in welchem Maße einen Erfolgssindikator beeinflussen. Darauf aufbauend kann ein aggregiertes Qualitätsmaß berechnet werden. Das Problem besteht darin, wie das Qualitätsmaß mit dem

Quantitätsmaß zu verbinden ist, d.h. wie die Adjustierung konkret umzusetzen ist. Dies hängt vom Einzelfall ab und wird von der Datenlage sowohl der Quantitätsindikatoren als auch der Qualitätsindikatoren bestimmt. In Kapitel 5 werden wir einige Möglichkeiten am Beispiel des Bildungssektors illustrieren.

3.6 Schlussfolgerungen

In diesem Kapitel wurde die Problematik der Outputmessung des öffentlichen Sektors, die sich aufgrund des Fehlens von Marktpreisen ergibt, diskutiert und Verfahren zur Lösung dieser Problematik vorgestellt. Es wurde gezeigt, dass die traditionelle Methode der Erfassung des staatlichen Outputs mittels Inputfaktoren schwerwiegende Mängel aufweist, da eine Produktivitätsmessung nicht möglich ist. Aus diesem Grund sind Verfahren der direkten Outputmessung zu bevorzugen. Die Implementierung dieser Verfahren ist jedoch auch mit einer Reihe von Problemen behaftet. Liegen geeignete Aktivitätsindikatoren vor? Sind diese Indikatoren repräsentativ für den Gesamtoutput eines Sektors? Muss auf Kostengewichte zur Aggregierung zurückgegriffen werden? Gibt es Daten zur Qualität der öffentlichen Leistungen? Die Antworten auf diese Fragen bestimmen die Ergebnisse und Güte der Outputmessung.

Wie wir im folgenden Kapitel anhand des Beispiels des Bildungssektors zeigen werden, haben sich bisher noch keine internationalen Standards herausgebildet. Wir zeigen, dass einige ausgewählte Länder sehr unterschiedliche Verfahren zur Outputmessung des öffentlichen Sektors verwenden. Des Weiteren analysieren wir anhand von stilisierten Berechnungen wie sich unterschiedliche Verfahren auf das Outputmaß auswirken können. Die Ergebnisse ermöglichen uns Rückschlüsse auf die Sensitivität des Outputmaßes. Führen unterschiedliche Verfahren zu sehr unterschiedlichen Ergebnissen wird zum einen die internationale Vergleichbarkeit des öffentlichen Outputs in Frage gestellt und zum anderen würde das Zweifel an der Verlässlichkeit von Produktivitätskennziffern des öffentlichen Sektors aufwerfen.

4. Statistische Übersicht zum öffentlichen Dienst im internationalen Vergleich

Wie im vorangegangenen Kapitel verdeutlicht ist die Messung der staatlichen Leistungserstellung mit großen Problemen behaftet. Aus diesem Grund wird in internationalen Vergleichen des Staatssektors üblicherweise auf Inputindikatoren zurückgegriffen. Insbesondere wird auf die Beschäftigung im öffentlichen Sektor und die mit den damit verbunden Ausgaben wie dem staatlichen Konsum rekurriert. Im Folgenden werden einige der wichtigsten statistischen Vergleichszahlen des Inputs, für die internationale Daten vorliegen, aufgeführt und diskutiert. Zusätzlich stellen wir einem qualitativen Indikator für das Leistungsangebot vor, welcher von der Weltbank veröffentlicht wird.

Tabelle 4.1 zeigt die Beschäftigung im staatlichen Sektor. Demnach ist in Deutschland der Anteil der Beschäftigten im Staatssektor an der Gesamtzahl der abhängig Beschäftigten seit 1980 kontinuierlich zurückgegangen und liegt im Jahre 2006 mit 11.9% nur über dem von Japan und Luxemburg. In den skandinavischen Ländern (Dänemark, Finnland, Norwegen und Schweden) ist dieser Anteil traditionell am Höchsten und liegt aktuell bei etwa 30%. In England ist der Anteil zwischen 1980 und 2000 um fast 10% zurückgegangen (von 30.3% auf 20.7%), ist seit dem jedoch stabil geblieben und sogar wieder leicht angestiegen.

Tabelle 4.2 zeigt die Arbeitnehmerentgelte des Staates. Die vom Staat gezahlten Arbeitnehmerentgelte als Anteil am Bruttoinlandsprodukt belaufen sich für die hier aufgeführten 26 EU-Mitgliedsstaaten seit 1999 im Durchschnitt auf etwa 11 %. Dieser Anteil blieb über den gesamten Beobachtungszeitraum hinweg relativ stabil. Betrachtet man insbesondere die Anteile der Arbeitnehmerentgelte am Bruttoinlandsprodukt für Deutschland, so ist zu konstatieren, dass Deutschland im hier betrachten Zeitraum stets unterhalb des EU-Durchschnitts liegt. Mit einem Anteil von etwa 8 % ist Deutschland ähnlich einzuordnen wie Luxemburg, Slowakei und Tschechische Republik, und gehört somit zu den Ländern mit dem niedrigsten Anteil. Zudem zeigt sich, dass der BIP-Anteil der Arbeitnehmerentgelte des Staates für Deutschland seit 1999 rückläufig ist. Zu den EU-Spitzenreitern im Jahre 2005 zählen die Länder Dänemark und Schweden mit einem Anteil von über 15 %.

Tabelle 4.1: Anteil der Beschäftigten im Staatssektor an der Gesamtzahl der abhängig Beschäftigten, 1980-2006, in Prozent

	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005	2006
Belgien	22.7	24.8	23.7	22.9	21.8	21.6	22.1	22.4	22.4	22.3	22.1
Tschechien	n.a.	n.a.	5.8	19.2	19.0	19.2	19.0	19.6	19.4	18.8	18.6
Dänemark	32.1	32.8	32.4	33.3	32.6	32.4	32.7	32.7	32.4	32.1	31.8
Finnland	23.3	25.8	27.2	31.0	28.5	28.4	28.7	28.9	29.0	28.9	28.7
Frankreich	23.1	25.6	25.2	26.3	24.3	24.1	24.2	24.4	24.5	24.5	24.5
Deutschland	**16.3**	**16.9**	**15.9**	**14.3**	**13.1**	**12.4**	**12.4**	**12.4**	**12.4**	**12.1**	**11.9**
Griechenland	19.4	21.6	23.5	22.6	23.7	21.8	22.3	21.9	21.6	n.a.	n.a.
Ungarn	n.a.	n.a.	n.a.	29.4	24.5	24.1	24.9	24.4	25.2	24.8	24.4
Irland	21.2	22.3	20.1	18.7	15.6	15.8	16.1	16.2	16.2	16.7	16.7
Italien	19.6	21.7	21.8	22.5	21.0	21.1	20.8	20.8	20.8	20.6	20.3
Luxemburg	15.4	15.4	13.5	12.7	11.4	11.0	10.8	10.9	10.9	10.7	10.6
Niederlande	15.0	17.4	15.4	13.9	12.6	12.7	13.0	13.4	13.6	13.5	13.5
Polen	n.a.	n.a.	n.a.	24.7	23.4	23.9	24.1	23.6	22.7	21.7	20.5
Portugal	15.1	18.1	19.4	23.7	24.4	25.8	26.4	25.0	24.7	24.5	24.2
Slowakei	n.a.	n.a.	n.a.	21.9	22.0	21.7	21.6	20.8	n.a.	n.a.	n.a.
Spanien	12.2	15.9	16.4	17.7	16.6	16.7	16.8	17.1	17.3	17.4	17.5
Schweden	35.4	37.4	35.8	36.2	33.1	32.7	33.0	33.3	33.6	33.4	33.4
Vereinigtes Königreich	30.3	28.5	24.9	22.6	20.6	20.7	21.0	21.6	21.9	22.1	22.1
Norwegen	28.4	29.7	32.9	35.4	34.0	34.5	34.4	34.6	34.6	34.4	33.9
Kanada	22.7	24.6	24.9	25.5	24.2	24.3	24.5	24.7	24.8	24.9	24.9
Japan	12.3	11.7	10.5	10.2	10.1	10.1	10.2	10.4	10.6	10.3	10.1
Vereinigte Staaten	17.7	16.6	16.6	16.3	15.6	15.9	16.3	16.4	16.3	16.2	16.1

Quelle: OECD, Economic Outlook Database, Paris 2006; Berechnungen von ifo.

Tabelle 4.2: Arbeitnehmerentgelte des Staates in Prozent des BIP (EU-27[1])

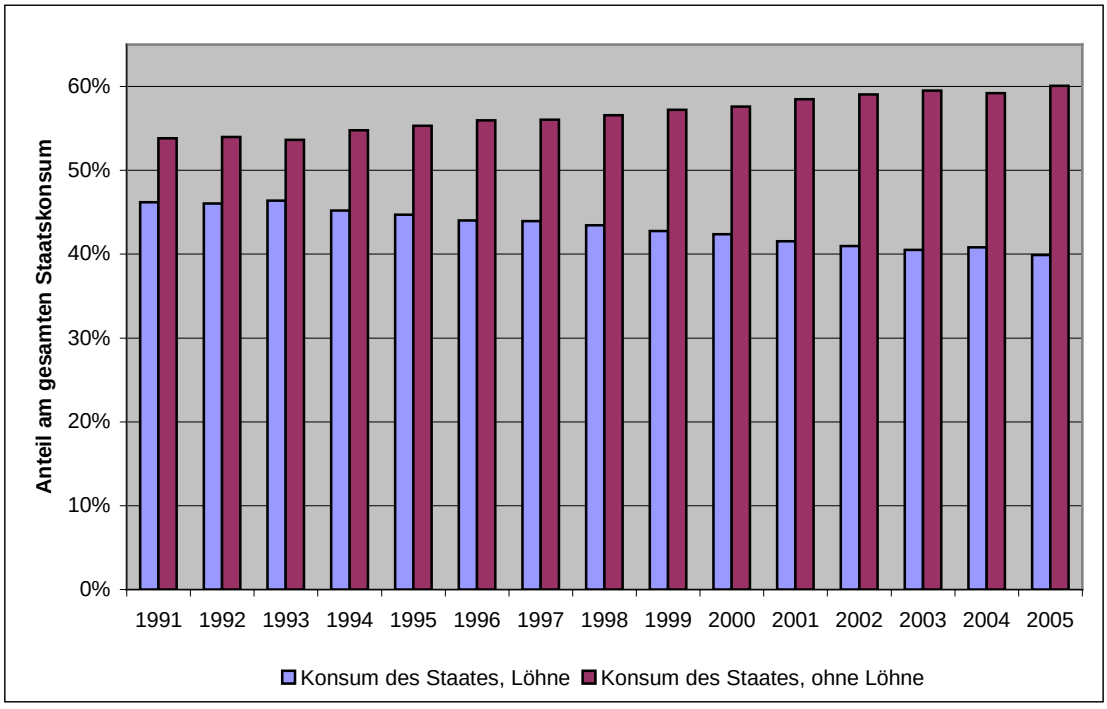
	1999	2000	2001	2002	2003	2004	2005
Belgien	11.8	11.5	11.7	12.2	12.3	12.0	12.1
Dänemark	17.5	17.1	17.4	17.8	18.0	17.9	17.2
Deutschland[2]	**8.2**	**8.1**	**7.9**	**7.9**	**7.8**	**7.7**	**7.5**
Estland	…	…	10.3	9.9	10.0	10.0	9.4
Finnland	13.5	13.0	12.9	13.2	13.6	13.5	13.8
Frankreich	13.5	13.3	13.3	13.5	13.5	13.3	13.2
Griechenland	11.7	11.5	11.4	12.1	11.8	12.5	12.1
Irland	8.1	8.0	8.3	8.6	9.0	9.4	9.5
Italien	10.6	10.4	10.5	10.6	10.8	10.8	10.9
Lettland	…	10.8	10.2	10.5	10.7	10.5	10.0
Litauen	…	…	…	11.4	10.9	10.9	10.4
Luxemburg (Grand-Duché)	7.8	7.5	7.9	8.0	8.1	8.2	8.1
Malta	13.8	13.0	14.9	14.5	14.7	14.7	14.0
Niederlande	9.7	9.5	9.6	9.8	10.1	10.0	9.9
Österreich	11.2	10.9	9.7	9.6	9.5	9.3	9.3
Polen	…	…	…	10.8	10.7	10.1	10.1
Portugal	13.7	14.2	14.3	14.7	14.1	14.1	14.3
Rumänien	…	…	…	8.3	8.3	8.1	9.0
Schweden	15.7	15.6	15.9	16.1	16.5	16.3	16.1
Slowakei	…	…	…	…	8.9	8.7	7.3
Slowenien	…	11.6	12.2	12.0	12.1	12.0	12.0
Spanien	10.5	10.3	10.1	10.0	10.1	10.1	10.0
Tschechische Republik	7.3	7.1	7.4	7.8	8.3	8.0	8.0
Ungarn	…	…	11.1	12.2	13.1	12.6	12.6
Vereinigtes Königreich	9.9	10.0	10.4	10.6	10.9	11.1	11.3
Zypern	13.5	13.5	13.2	13.8	15.6	14.8	14.8
Durchschnitt	11.6	11.3	11.4	11.4	11.5	11.4	11.3

Quelle: Eurostat Datenbank, Sektor Staat, 2007.

Anmerkungen: [1]EU-27 enthalten die Mitgliedsstaaten der Europäischen Union mit Ausnahme von Bulgarien (geringe Datenverfügbarkeit). [2]Deutschland (einschließlich ex-DDR). Methodischer Rahmen zur Abgrenzung des Staates ist das Europäische System Volkswirtschaftlicher Gesamtrechnungen, Ausgabe 1995, (ESVG 95) und die "Klassifikation der Ausgaben des Staates nach dem Verwendungszweck" (COFOG), Vereinte Nationen, 1999. Der Staat umfasst dabei die Verwendungszwecke: Allg. öffentliche Dienstleistungen, Verteidigung, Öffentl. Ordnung und Sicherheit, Wirtschaftl. Angelegenheiten, Umweltschutz, Wohnungs- und Gemeindewesen, Gesundheit, Erholung, Kultur und Religion, Erziehung, Soziale Sicherung.

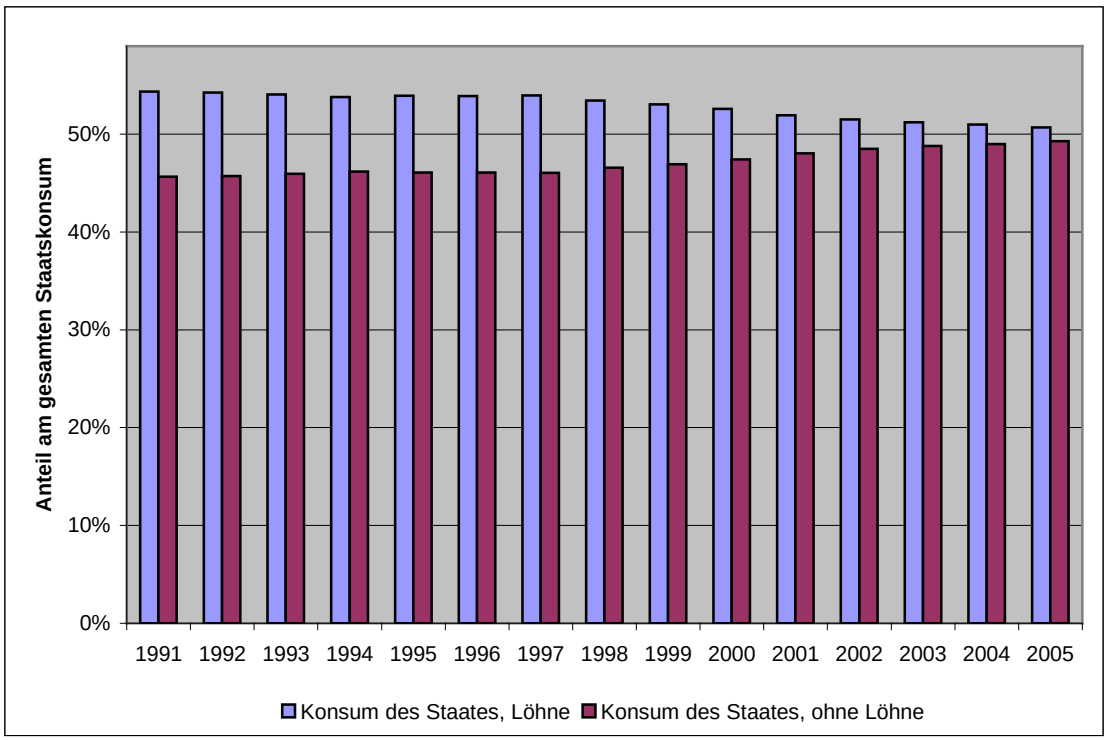
Wie Abbildung 4.1. zeigt, ist im Hinblick auf den staatlichen Verbrauch insgesamt zu konstatieren, dass sich die staatlichen Konsumausgaben in Deutschland bereits seit der Wiedervereinigung stärker auf sonstige, d.h. auf nicht lohnbezogene Ausgaben, als auf Löhne verteilen. Zudem erfolgte seit 1991 eine deutliche Reduzierung des Anteils der auf Löhne entfallenden Staatsausgaben. Lag der Lohnanteil an den Konsumausgaben 1991 noch bei ca. 45 %, belief er sich 2005 nur noch auf 40 %. Die Umschichtung innerhalb des Staatsverbrauchs führte zu einem Anstieg des sonstigen Konsumanteils von 55 % in 1991 auf nahezu 60 % in 2005. Eine Reduktion des Anteils der Lohnausgaben am Staatskonsum lässt sich wie Abbildung 4.2 zeigt für die gesamten EU12-Mitgliedstaaten erkennen, wobei auf dem aggregierten Niveau eher von einer Angleichung als einer Divergenz der Anteile, wie im Falle Deutschlands, zu sprechen ist.

Abbildung 4.1: Konsum des Staates aufgeteilt nach Löhnen und sonstigen Konsumgütern, Anteil am staatlichen Gesamtkonsum in Prozent, Deutschland, 1991–2005



Quelle: OECD, Economic Outlook, 2007.

Abbildung 4.2: Konsum des Staates aufgeteilt nach Löhnen und sonstigen Konsumgütern, Anteil am staatlichen Gesamtkonsum, EU12 Mitgliedsstaaten, 1991–2005

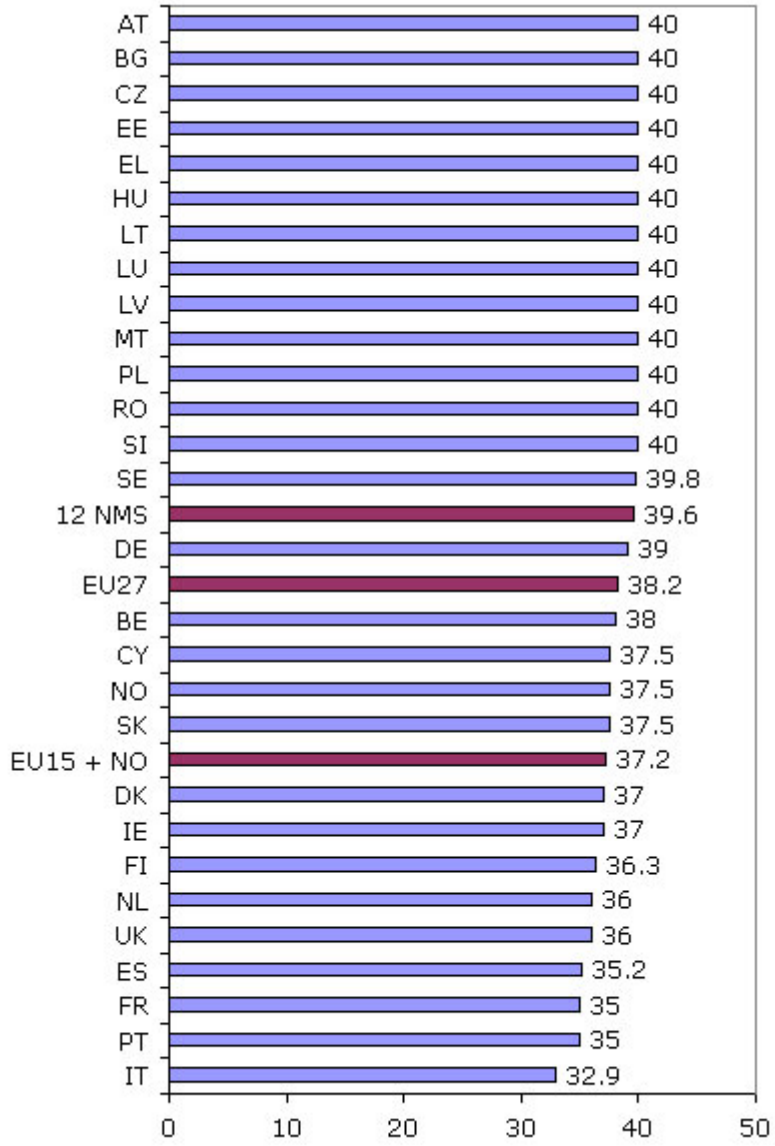


Quelle: OECD, Economic Outlook, 2007.

Anmerkungen: EU12 umfasst die EU-Mitgliedsstaaten: Deutschland, Frankreich, Italien, Österreich, Belgien, Finnland, Irland, Luxemburg, Niederlande, Portugal, Spanien und Griechenland.

Ein Überblick der tariflichen Arbeitszeit wird in Abbildung 4.3. gegeben. Hinsichtlich der tariflich vereinbarten wöchentlichen Arbeitsstunden für Behördenbedienstete liegt Deutschland mit 39 Stunden pro Woche knapp über dem Durchschnitt der 27 EU Mitgliedsländer. Mit 40 Stunden etwas länger als in Deutschland, wird lediglich in Österreich, Luxemburg, Griechenland sowie den meisten osteuropäischen Länder gearbeitet. Am unteren Ende der Skala finden sich Länder wie Spanien, Frankreich und Portugal. Mit Abstand am Geringsten ist die Wochenarbeitszeit in Italien mit knapp 33 Stunden.

Abbildung 4.3: Arbeitszeit der Behördenbediensteten, Tarifvereinbarungen wöchentlicher Arbeitsstunden, 2006



Quelle: European Industrial Relations Observatory, EIRO, Working Time Developments 2006.

Anmerkungen: Zahlen für Deutschland beziehen sich auf Gesamtdeutschland, Quelle: Archiv Tarifvereinbarungen des Wirtschafts- und Sozialwissenschaftlichen Instituts (WSI) der Hans-Böckler-Stiftung. Tariflich vereinbarte wöchentliche Arbeitsstunden vor 2006: 38,5 für West-Deutschland, 40 für Ost-Deutschland. NMS steht für neue EU-Mitgliedsstaaten. Viele der Zahlen beziehen sich auf tarifliche Vereinbarungen in den jeweiligen Ländern, während in einigen Ländern Arbeitszeiten gesetzlich vorgeschrieben werden. Dies trifft insbesondere für die neuen EU-Mitgliedsstaaten zu.

Die hier dargelegten Indikatoren zeigen, dass Deutschland im internationalen Vergleich sowohl was die Anzahl der Beschäftigten als auch die damit verbundenen Kosten im Staatssektor betrifft am unteren Rand liegt. Der Beschäftigtenanteil im Staatssektor belief sich mit etwa 12 % der abhängigen Beschäftigten in 2006 auf einem der niedrigsten Niveaus in der OECD.

Hinsichtlich der Kostenentwicklung bei den Beschäftigten befindet sich Deutschland stets unterhalb des EU-Durchschnitts, wobei ein kontinuierlicher Abbau der Kosten zu konstatieren ist. Ebenso ist der Anteil der Entlohnung an den gesamten Konsumausgaben des Staates seit der Wiedervereinigung durchgehend von etwa 45 % in 1991 auf 40 Prozent in 2005 zurückgegangen. Bei den tariflich vereinbarten wöchentlichen Arbeitsstunden liegt Deutschland mit 39 Stunden pro Woche in etwa in Höhe des EU-Durchschnitts mit 38.2 Wochenarbeitsstunden.

Was die Leistungsseite angeht ist ein Vergleich notorisch schwierig, wie die obige Analyse zeigt. Die Outputmessung wird gegenwärtig noch entwickelt und liegt allenfalls für Teilbereiche vor. Was verfügbar ist, sind verschiedene qualitative Indikatoren. Als Beispiel sei hier auf einen von der Weltbank konzipierten Indikator eingegangen. Dieser Indikator der *Government Effectiveness*, der die Effektivität und Qualität des Staatssektors insgesamt auf Grundlage von Experteneinschätzungen und Umfragen misst, weist für Deutschland neben einem Rückgang der relativen Effektivität im Zeitablauf (1996-2004) auch eine niedrige Platzierung im OECD-Vergleich aus (vgl. Tabelle 4.3). Der Indikator Government Effectiveness misst dabei nicht nur die Qualität des öffentlichen Dienstes und der Bürokratie in der Einschätzung der Befragten, sondern auch die Unabhängigkeit des öffentlichen Dienstes von politischem Druck sowie die Qualität von Politikmaßnahmen hinsichtlich ihrer Ausgestaltung (Konzeption), Umsetzung und Glaubwürdigkeit.

Tabelle 4.3: Effektivität des Staatssektors (Government Effectiveness), 1996–2006

Länder	1996	1998	2000	2002	2003	2004	2005	2006
Australien	1.43	1.89	1.83	1.81	1.97	2.06	1.89	1.94
Österreich	2.04	1.77	1.92	1.97	1.94	1.78	1.62	1.62
Belgien	1.94	1.67	1.7	2.03	1.91	1.79	1.66	1.64
Kanada	2.07	1.86	1.91	2.09	2.12	2.05	1.94	2.03
Tschechien	0.84	0.72	0.77	0.87	0.75	0.75	1.01	1.01
Dänemark	2.18	1.94	1.96	2.14	2.22	2.26	2.14	2.29
Finnland	2.1	1.93	2.01	2.18	2.23	2.09	2.09	2.08
Frankreich	1.76	1.36	1.6	1.59	1.57	1.49	1.47	1.2
Deutschland	**2.07**	**1.81**	**1.91**	**1.79**	**1.48**	**1.43**	**1.51**	**1.52**
Griechenland	0.79	0.87	0.73	0.89	0.84	0.81	0.66	0.62
Ungarn	0.59	0.92	0.94	1.02	0.86	0.82	0.75	0.71
Island	1.5	2.07	2.09	2.01	2.23	2.23	2.22	2.13
Irland	1.84	1.72	1.78	1.75	1.62	1.58	1.64	1.53
Italien	0.97	0.93	0.86	0.88	0.88	0.68	0.6	0.38
Japan	1.34	1.03	1.05	1.01	1.09	1.11	1.16	1.29
Korea, Süd	0.92	0.41	0.77	0.95	0.9	0.93	1.01	1.05
Luxemburg	2.27	2.05	2.06	2.18	2.09	2.12	1.95	1.73
Mexiko	-0.04	0.34	0.31	0.31	0.18	0.13	0.01	0.16
Niederlande	2.24	2.12	2.09	2.11	2.06	2.09	1.96	1.86
Neuseeland	2.22	1.77	1.63	1.8	2.01	2.16	1.91	1.94
Norwegen	2.22	2.04	1.92	2	2.07	2.13	2.01	2.1
Polen	0.78	0.73	0.59	0.56	0.55	0.45	0.54	0.49
Portugal	1.1	1.37	1.1	1.19	1.24	1.07	1.03	0.95
Slowakei	0.54	0.25	0.3	0.45	0.57	0.75	0.95	0.91
Spanien	1.6	1.7	1.75	1.82	1.76	1.36	1.4	1.05
Schweden	2.12	2.05	2	2.07	2.09	2.07	1.95	2
Schweiz	2.39	1.99	2.16	2.19	2.05	2.3	2.05	2.13
Türkei	-0.18	-0.22	-0.06	0	0.11	0.07	0.21	0.23
Ver. Königreich	2.09	2.04	1.9	1.97	1.94	1.92	1.71	1.83
Ver. Staaten	2.12	1.65	1.88	1.78	1.77	1.81	1.6	1.64

Quelle: Weltbank, Governance Matters Datenbank, 2007.

Anmerkungen: Die Indexwerte sind skaliert von -2.5 bis +2.5, wobei höhere Werte bessere Government Effectiveness bedeuteten. Ein Wert von 0 entspricht dem Mittelwert über **alle** erfassten Länder (212 in 2006).

Die Weltbank versucht in ihren Governance Matters Indikatoren, abstrakte (nicht direkt messbare) Konzepte durch eine Vielzahl von (messbaren) Teilaspekten zu erfassen. Dazu werden statistische Verfahren herangezogen. Die Indizes bilden sich aus der Aggregation von Einschätzungen einer großen Anzahl von Unternehmen, Bürgern und Experten. Die Daten werden von Umfrageinstituten, Think-Tanks, Nicht-Regierungsorganisationen und internationalen Organisationen erhoben. Für Deutschland werden beispielsweise für den Government Effectiveness Indikator für das Jahr 2006 Informationen aus 10 verschiedenen Quellen herangezogen.[16] Es ist indessen darauf zu achten, dass der Indikator auf die Leistung des öffentli-

[16] Diese Quellen umfassen Business Environment Risk Intelligence (BERI) Business Risk Service, Economist Intelligence Unit, Gallup World Poll, Global E-Governance Index, Global Insight (GI) Business Conditions and Risk Indicators, Global Insight (GI) Global Risk Service, Institute for Management and Development (IMD)

chen Sektors insgesamt abzielt, ohne dass dabei zwischen der Produktivität im öffentlichen Dienst also dem Umfang der Leistungen pro Beschäftigten und dem Umfang an öffentlichen Leistungen insgesamt unterschieden wird.

Im Zeitablauf hat sich die Einschätzung der Effektivität des Staatssektors in Deutschland bis 2004 kontinuierlich verschlechtert. Danach ist eine leichte Verbesserung zu verzeichnen. Aufgrund der Konstruktion dieses Indikators können diese Veränderungen im Zeitablauf durch zwei Effekte begründet sein: Erstens können sich die Einschätzungen über die Effektivität des Staatssektors im Durchschnitt *aller* Länder verschlechtert haben. Zweitens kann dies auch Ausdruck einer *relativen* Verschlechterung Deutschlands im internationalen Vergleich sein. Verglichen mit anderen OECD Ländern lag Deutschland 2006 nur im unteren Mittelfeld. Bei den Indexwerten handelt es sich um Schätzwerte, die mit einer gewissen Unsicherheit behaftet sind (der Grad der Unsicherheit wird von der Weltbank durch die Angabe von Standardabweichungen dargestellt. Die Standardabweichungen sind in dieser Tabelle nicht angegeben). Insofern sollte von einer Interpretation der konkreten Länderplatzierung abgesehen werden.

Generell ist aber zu betonen, dass erhebliche methodische Einwände gegen entsprechende Indikatoren vorgebracht werden können. Dies beginnt schon bei den zugrunde liegenden Befragungsdaten, die anders als bei sozialstatistischen Datensätzen nicht repräsentativ erhoben werden. Man behilft sich hier mit Umfragen unter „Experten“, einem kleinen Kreis nicht zufällig ausgewählter Beobachter. Ein fundamentales Problem ist auch die Gewichtung der verschiedenen Indikatoren. Im Prinzip werden willkürliche Gewichtungen verwendet, durch die Indikatoren letztendlich völlig beliebige Werte liefern können. Der Aussagewert ist daher äußerst fragwürdig. Leistungsindikatoren dieser Form sind deswegen kein Substitut für eine sorgfältige Outputmessung. Aus diesem Grund wenden wir uns im Folgenden wieder der Outputmessung zu. Das nachfolgende Kapitel beschäftigt sich exemplarisch mit der Outputmessung im Bildungsbereich im internationalen Vergleich.

World Competitiveness Yearbook, Merchant International Group Gray Area Dynamics, Political Risk Services International Country Risk Guide (ICRG), World Economic Forum (WEF) Global Competitiveness Report.

5. Illustration der Outputmessung anhand des Bildungssektors

5.1 Einleitung

Die Ausführungen des 3. Kapitels haben gezeigt, dass Methoden der direkten Outputmessung aus theoretischer Sicht den traditionellen Inputverfahren zu bevorzugen sind. Die konkrete Umsetzung der direkten Outputverfahren und insbesondere die Möglichkeit der Qualitätsadjustierung hängen jedoch in entscheidendem Maße von den verfügbaren Daten ab. Beispielsweise bestimmt die Disaggregationstiefe der Daten die Erfassbarkeit von Qualitätsstruktureffekten. Ebenso sind Informationen zu Outcome-Maßen unerlässlich für eine direkte Qualitätsadjustierung. Stehen Outcome-Maße zur Verfügung, besteht das Problem der Auswahl der am besten geeigneten Indikatoren sowie deren konkrete Berücksichtigung in der Indexkonstruktion. Datenrestriktionen sowie das Fehlen von konkreten international verpflichtenden Standards zur Qualitätsadjustierung haben dazu geführt, dass die statistischen Ämter in verschiedenen Ländern unterschiedliche Wege der Outputmessung des öffentlichen Sektors beschritten haben. In diesem Kapitel wollen wir uns ausführlicher mit der praktischen Umsetzung der direkten Outputmessung beschäftigen. Dazu soll zunächst ein internationaler Überblick über verschiedene Implementierungsansätze gegeben werden. Daran anschließend untersuchen wir anhand von stilisierten Berechnungen die Sensitivität verschiedener Ansätze auf die Ergebnisse bezüglich Output und Produktivität.

Die Ausführungen dieses Kapitels beschränken sich dabei auf einen Bereich des öffentliche Sektors: den Bildungsbereich. Dies geschieht aus folgendem Grund: Der Bildungsbereich ist überwiegend durch individuelle Dienstleistungen gekennzeichnet. Wie im 3. Kapitel diskutiert, lassen sich die Verfahren der direkten Outputmessung für individuelle Güter am besten umsetzen. Des Weiteren ist in diesem Bereich eine Vielzahl von notwendigen Daten vorhanden. Daher lassen sich die verschiedenen Ansätze für diesen Bereich besonders gut illustrieren. Weitere Sektoren mit überwiegend staatlicher Aktivität sind der Gesundheitsbereich sowie der Bereich der Öffentliche Verwaltung, Sozialversicherungen und Verteidigung. In Deutschland ist der Gesundheitsbereich zum Großteil durch Marktproduktion gekennzeichnet. Da in diesem Fall Preise vorliegen, kann die Outputmessung wie bei Marktproduktion üblich über Deflationierungsverfahren erfolgen. Dies ist auch das Verfahren, das das Statistische

Bundesamt in Deutschland verwendet.[17] Der Bereich Öffentliche Verwaltung, Sozialversicherungen und Dienstleistungen ist demgegenüber überwiegend durch kollektive Leistungen gekennzeichnet. Auf Grund der Schwierigkeit der Outputdefinition und den sich daraus ergebenen Problemen der Anwendbarkeit direkter Outputverfahren bei kollektiven Leistungen wird in Deutschland sowie in den meisten anderen Ländern in diesem Bereich weiterhin das Inputverfahren verwendet.[18]

5.2 Internationaler Methodenüberblick der Konstruktion von Output-Indizes

Zur Gegenüberstellung internationaler Methoden bei der Konstruktion von Outputindizes im Bildungsbereich werden im Folgenden einzelne Indexkomponenten separat aufgeführt. Dabei erfolgt eine Aufteilung nach der Disaggregationstiefe (Stratifikation), der jeweils verwendeten Quantitätsmaße, ihrer entsprechenden Gewichte und der dazu durchgeführten Qualitätsadjustierung (sofern vorhanden). Die Gegenüberstellung der einzelnen Komponenten erfolgt für die Länder Deutschland, Finnland, Italien, Schweden und das Vereinigte Königreich. Die Auswahl der Länder resultiert in erster Linie aus der Verfügbarkeit der notwendigen Informationen, die für methodische Vergleiche internationaler Outputindizes benötigt werden. Außerdem soll die Gegenüberstellung einen Überblick über die Vielfalt der unterschiedlichen internationalen Ansätze liefern. Die folgenden Ausführungen beschränken sich auf die wichtigsten Charakteristika und Innovationen der jeweiligen Outputmessung. Eine detaillierte Übersicht der einzelnen Verfahren findet sich in Tabelle 5.1.

Deutschland

Mit der Revision der Volkswirtschaftlichen Gesamtrechnung des deutschen Statistischen Bundesamtes des Jahres 2005 wurde die Outputmessung des Bildungssektors vom traditionellen Inputverfahren auf direkte Outputmessverfahren umgestellt. Das deutsche System zeichnet sich durch eine hohe Disaggregationstiefe aus. Der Bereich der Bildung wird in die drei Sektoren Staat, Organisationen ohne Erwerbszweck und Unternehmen untergliedert, wobei jeder einzelne Sektor wiederum eine Segmentierung nach zehn Bildungssektoren erfährt. Im universitären Bereich und den Fachhochschulen wird eine weitere Untergliederung nach neun

[17] vgl. Braakmann et al., 2005.
[18] Der Atkinson Report diskutiert eine Reihe von Möglichkeiten, wie auch in diesem Bereich direkte Outputindikatoren verwendet werden können (Atkinson, 2005, S. 141ff)

Fächergruppen vorgenommen, während im Bereich der Kindergärten nach Halb- und Ganztagseinrichtungen unterschieden wird. Zur Bestimmung der Mengenkomponenten zieht das Statistische Bundesamt für die unteren Bildungssektoren die Unterrichtsstunden, für die Universitäten und Fachhochschulen die Anzahl der Studierenden in den jeweiligen Fächergruppen und im Bereich Kindergarten die Betreuungsstunden heran. Zur Aggregation der entsprechenden Mengenkomponenten werden dann Kostengewichte in Form von Durchschnittskosten eingesetzt. Wie bereits in den vorherigen Kapiteln erwähnt, können (Durchschnitts-) Kostengewichte unter bestimmten Annahmen zur Approximation der Zahlungsbereitschaft der Konsumenten verwendet werden. Verzerrungen treten jedoch auf, wenn die Kosten langfristig nicht die Zahlungsbereitschaft widerspiegeln oder wenn die verwendeten Durchschnittskosten von den marginalen Kosten abweichen. Durch die hohe Disaggregationstiefe der einzelnen Mengenkomponenten und deren Aggregation mittels Kostenanteilen wird eine Berücksichtigung von Qualitätsstruktureffekten realisiert. Diese induzieren eine Veränderung der Bildungsqualität rein aufgrund von Anteilsverschiebungen von Schülern/Studenten zwischen Schultypen/Studiengängen mit unterschiedlichen zurechenbaren Kosten pro Schüler/Student sofern die Durchschnittskosten Qualitätsunterschieden entsprechen. Eine explizite Qualitätsadjustierung wird für den deutschen Bildungssektor nicht durchgeführt. Während das Fehlen einer expliziten Berücksichtigung von Qualitätsveränderungen als Mangel aufgefasst werden kann, stellt die Umstellung auf ein direktes Outputverfahren sowie die tiefe Disaggregationsebene eine eindeutige Verbesserung gegenüber dem Inputverfahren dar.[19]

Nach Informationen des Statistischen Bundesamtes hat die für Deutschland eingeführte Umstellung von der Input- auf die direkte Outputmethode zu einer Erhöhung des Outputvolumens im Bildungsbereich um 0.7 Prozentpunkte über den Zeitraum 1995-2003 geführt. Einem Anstieg von 3.8% nach der Inputmethode steht ein Anstieg um 4.5% nach der direkten Outputmethode zwischen 1995 und 2003 gegenüber.[20]

[19] Weitere Informationen zur deutschen Methode können in einem Projektbericht des Statistischen Bundesamtes aus dem Jahr 2004 gefunden werden: „Weiterentwicklung der Methoden der Preis- und Volumenmessung in der Volkswirtschaftlichen Gesamtrechnung“ (http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Infos/Weiterentwicklung__VGR,property=file.pdf).

[20] Vgl. Braakmann et al., 2005.

Finnland

Die Konstruktion des Bildungsoutputs in Finnland folgt ebenso wie in Deutschland zunächst einmal einer Untergliederung des Bildungssektors nach einzelnen Bildungsstufen. Neben einer Unterteilung hinsichtlich zentraler und regionaler Zuständigkeiten und nach Schülertypen (reguläre und Sonderschüler), erfolgt auch im universitären Bereich eine Disaggregation nach Fächergruppen (siehe Tab. 5.1). Als Mengenmaß wird im nicht-universitären Bereich die Anzahl der Schüler verwendet. Die Verwendung der Schülerzahl als Quantitätskomponente muss gegenüber den Unterrichtstunden im deutschen System nur als second best Lösung angesehen werden. Die Innovation des finnischen Systems besteht in der Benutzung von erworbenen Kreditpunkten als Mengenkomponente im universitären Bereich (für weitere Ausführungen zu den Mengenkomponenten anderer akademischer Ausbildungsstufen siehe Tab. 5.1). Erworbene Kreditpunkte sind ein Beispiel für die Verwendung von Erfolgs- oder Outcome-Indikatoren als Quantitätsmaß. Kreditpunkte bilden neben einer reinen Mengen- auch eine Qualitätskomponente ab, da eine im Zeitablauf steigende Zahl an Kreditpunkten pro Student als Verbesserung der Bildungsleistung interpretiert werden kann. Dieser Versuch, die Qualität des Universitätssystems mittels Erfolgsmaßen direkt zu erfassen, kann als Vorteil gegenüber dem deutschen Verfahren angesehen werden.

Vorläufige Berechnungen des Statistikamtes in Finnland für den Zeitraum 2000-2005 zeigen, dass bei polytechnischen Schulen die Verwendung der Zahl der erworbenen Kreditpunkte anstelle der Anzahl der Studenten nur zu marginalen Unterschieden geführt hat. Dagegen hat die Einführung von Kreditpunkten anstelle von Studentenzahlen im universitären Bereich zu einem signifikant höheren Output geführt (der Unterschied in 2005 beträgt ca. 4 Prozentpunkte).[21]

Italien

Auch in Italien wird versucht, die Qualität des Bildungssystems explizit zu erfassen. Anders als beim finnischen Verfahren, wird jedoch die Quantität weiterhin über Aktivitätsindikatoren erfasst. Im schulischen und akademischen Bereich werden die Anzahl der Schüler bzw. der Studenten herangezogen. Jedoch werden diese Aktivitätsindikatoren basierend auf Qualitätsindikatoren herauf- bzw. herabskaliert. Für den schulischen Bereich wird die Klassengröße als Qualitätsindikator implementiert, wobei durch die Verwendung einer Kongestions-Funktion,

[21] Vgl. hierzu und zu weiteren Ausführungen zum finnischen Verfahren Heikkinen und Hautakangas, 2007.

welche Unterauslastungen oder Überfüllungskosten abbilden soll, die tatsächliche Schülerzahlen pro Klasse in „standardisierte“ Schülerzahlen pro Klasse transformiert werden. Für jede Bildungsstufe im Schulsystem wird eine minimale und maximale Klassengröße basierend auf Beschlüssen des Bildungsministeriums definiert. Überschreitet die Klassengröße die Minimalgröße, so nimmt die Funktion einen konkaven Verlauf an, d.h. ein zusätzlicher Schüler entspricht weniger als einem „standardisierten“ Schüler. Die tatsächliche Schülerzahl wird somit herabskaliert. Für das Universitätssystem werden dahingegen zwei Indikatoren konstruiert. Ein Indikator wird als Quotient aus der Zahl der Studenten, die die Regelstudienzeit nicht überschritten haben, zu allen Studenten gebildet. Der andere Indikator greift auf die Studienzeit zurück und stellt dabei die Regelstudienzeit zur tatsächlichen Studienzeit ins Verhältnis. Beide Indikatoren können somit einen Maximalwert von 1 erreichen. Bei der Konstruktion eines einzelnen Qualitätsindikators für das Universitätssystem gehen beide Qualitätsindikatoren gleichgewichtet ein. Der Qualitätsindikator wird dann mit der Studentenzahl multipliziert. Nimmt der Qualitätsindikator einen Wert kleiner 1 an (einer oder beide der Komponenten sind in diesem Fall unter ihrem Maximalwert) so impliziert dies wiederum eine Herabskalierung der tatsächlichen Studentenzahl. Im Vergleich zum finnischen System, in dem nur Qualitätsveränderungen im tertiären Bereich berücksichtigt werden, kann die zusätzliche Qualitätsadjustierung im schulischen Bereich durch die Berücksichtigung der Klassengröße als Verbesserung angesehen werden.[22]

Schweden

Ein weiteres Beispiel für eine explizite Qualitätsadjustierung im Schulsystem bildet das schwedische System. Diese Adjustierung erfolgt jedoch nur für die Abschlussklasse (neunte Bildungsstufe). Beim schwedischen Verfahren wird die Zahl der Schüler der neunten Klasse zunächst um diejenigen Schüler bereinigt, die keinen Abschluss erhalten haben. Es wird also wiederum wie beim finnischen System ein Outcome-Indikator statt eines reinen Aktivitätsindikators zur Messung der Quantität verwendet. Darauf aufbauend wird diese Zahl noch herauf- bzw. herabskaliert durch einen Faktor, der die Veränderung der Schüler mit Prädikatsexamen erfasst sowie einem zweiten Faktor, der die Veränderung des Prozentsatzes von Schülern berücksichtigt, die sich für eine Ausbildung in der nächst höheren Bildungsstufe qualifiziert haben. Das schwedische System ist somit ein Beispiel für die Kombination von einem Outcome-Maß als Mengenkomponente sowie der Verwendung von weiteren Outcome-Maßen

[22] Für eine detaillierte Darstellung des italienischen Systems vgl. Collesi et al., 2007.

zur Qualitätsadjustierung dieser Mengenkomponente. Negativ zu bewerten ist die Tatsache, dass die Berücksichtigung der Qualität ausschließlich für eine einzige Schulstufe (neunte Klasse) erfolgt.[23]

Vereinigtes Königreich

Als letztes Beispiel soll das Verfahren des Vereinigten Königreichs vorgestellt werden. Eine Besonderheit des britischen Systems besteht in der Verwendung von Anwesenheitszeiten zur Messung der Quantitätskomponente des schulischen Outputs. Anwesenheitszeiten haben gegenüber alternativen Mengenmaßen wie Schulstunden und insbesondere Anzahl der Schüler den Vorteil, dass sie die tatsächlich von den Schülern in Anspruch genommenen Unterricht messen. In den Bereichen, in denen eine Erhebung von Anwesenheitszeiten nur schwerlich umsetzbar ist, wie bspw. in Universitäten, wird nach wie vor auf Studentenzahlen zurückgegriffen. Die einzelnen Mengenkomponenten werden dann mittels Kostengewichten aggregiert. Eine besondere Art der Qualitätsadjustierung erfolgt für die gesetzlichen und weiterführenden Bildungsstufen, d.h. für primäre und sekundäre Schulstufen. Bei diesen Schulstufen wird eine jährliche Verbesserung der Qualität von 0.25% angesetzt. Diese geht auf historische Resultate für den britischen Abschlusstest GCSE (General Certificate of Secondary Education) zurück, wobei GCSE-Durchschnittswerte, die seit Mitte der 1990er erhoben werden, Verwendung finden.

An dieser Stelle sei noch kurz auf einen schottischen Vorschlag zur Messung des Schuloutputs eingegangen. Dieser Vorschlag sieht die Verwendung von Outcome basierten Gewichten zur Aggregation des Schuloutputs vor und kann somit als Verbesserung gegenüber den bisher ausschließlich verwendeten Kostengewichten angesehen werden. Es muss jedoch betont werden, dass es sich bisher lediglich um einen Vorschlag handelt, der noch nicht Eingang in die Outputmessung des Bildungssystems im Vereinigten Königreich gefunden hat. Zur Bestimmung des Index wird die Anzahl der Schüler der Abschlussklasse nach unterschiedlichen Qualifikationsstufen (hoch, mittel und niedrig) unterteilt, die unterschiedliche Bildungsabschlüsse reflektieren. Um die Mengenkomponenten der jeweiligen Stufen aggregieren zu können, werden Gewichte, die auf den zukünftig zu erwartenden Löhnen der Schüler basie-

[23] Eine Beispielrechnung des Statistikamtes Schwedens zeigt demnach auch nur eine marginale Erhöhung des Outputvolumens aller gesetzlich vorgeschriebenen Schulklassen nach Berücksichtigung der erwähnten Qualitätsadjustierungen für die neunte Klasse. Vgl. hierzu sowie zu weiteren Ausführungen der schwedischen Methodik Magnusson, 2007.

ren, verwendet. Die Berechnung der erwarteten Löhne nach Abschluss einer bestimmten Bildungsstufe erfolgt durch die Multiplikation der durchschnittlichen Löhne der jeweiligen Bildungsgruppe multipliziert mit der Wahrscheinlichkeit in Arbeit zu gelangen. Letzteres wird über die Multiplikation mit der für die jeweilige Bildungsstufe assoziierten Beschäftigungsquote umgesetzt. Den zu erwartenden Löhnen lässt sich dann über ihren relativen Anteil an der Summe aller Löhne ein entsprechendes Gewicht zuweisen, mit dem die Mengen schließlich aggregiert werden. Diese Art der Gewichtung stellt eine eindeutige Verbesserung gegenüber der zumeist verwendeten Kostengewichte dar, da die erwarteten Löhne in engerem Zusammenhang zur Zahlungsbereitschaft der Konsumenten stehen als Kosten. Insofern werden Qualitätsstruktureffekte, welche durch die Verschiebung zwischen Qualifikationsstufen entstehen, besser abgebildet.

Berechnungen des Office for National Statistics (ONS) in Großbritannien zeigen, dass die bisher in der volkswirtschaftlichen Gesamtrechnung implementierten Qualitätsadjustierungen, inkl. der revidierten Mengenindizes, eine Erhöhung des Outputindex um 2,0 Prozentpunkte über den Zeitraum 1996–2006 bewirkt haben (5,8 Prozent gegenüber den 3,8 Prozent ohne jegliche Qualitätsadjustierung). Würden alle bisher noch in der Testphase befindlichen Qualitätsadjustierungen (hier nicht besprochen) für den Bildungssektor umgesetzt, würde dies den aktuellen Output nach Berechnungen des ONS um weitere 21.3! Prozentpunkte innerhalb derselben Zeitperiode ansteigen lassen (der Gesamtanstieg des Bildungsoutputs beliefe sich somit auf 27.1% zwischen 1996 und 2006). Auch wenn diese Berechnungen auf bisher nur in der Testphase befindlichen Methoden beruhen, so zeigt diese Zahl dennoch das enorme Ausmaß, welches die Berücksichtigung der Qualität auf die Outputmessung haben kann.[24]

Die in diesem Abschnitt dargestellten internationalen Methoden zur Konstruktion von Outputindizes verdeutlichen die vielfältigen Vorgehensweisen und damit aber auch Schwierigkeiten bei der Festlegung eines internationalen Status quo. Explizite Qualitätsadjustierungen durch Outcome-Indikatoren verlangen nach einem deutlich umfangreicheren Datenmaterial, das möglicherweise nicht in allen Ländern in ausreichender Form vorhanden ist. Dennoch scheint sich neben der weit verbreiteten indirekten Adjustierung über Qualitätsstruktureffekte, wie sie in Deutschland ausschließlich praktiziert wird, verstärkt die Tendenz einer zusätzlichen expliziten Qualitätsadjustierung durch die Verwendung von Outcome-Indikatoren herauszubilden.

[24] Für detaillierte Informationen zum Status Quo im Vereinigten Königreich sowie zu den in der Testphase befindlichen Qualitätsadjustierungen siehe ONS, 2007.

Aufgrund der geringen Informationslage ist es bislang schwierig, abschließend zu beurteilen, wie sich der gemessene Bildungsoutput durch die Umstellung von Input- auf qualitätsadjustierte direkte Outputverfahren verändert. Insbesondere die Berechnungen Großbritanniens zeigen jedoch, dass die Vernachlässigung von Qualitätsveränderungen zu einer signifikanten Unterschätzung des Outputvolumens führen kann. Diese Tendenz lässt sich, wenn auch in geringerem Maße, ebenfalls für die anderen hier untersuchten Länder konstatieren.

Tab.: 5.1: Internationaler Methodenüberblick

Land	Disaggregationstiefe	Quantitätsmaße	Gewicht	Qualitätsadjustierung	Ergebnisse gegenüber alter Methode
Deutschland	Unterteilung in **3Sektoren**: Staat, Org. o. Erwerbszweck und Unternehmen, die jeweils untergliedert werden nach **10 Bildungssektore**n. Im Bereich Universitäten, und Fachhochschulen weitere Untergliederung nach **9 Fächergruppen** Im Bereich Kindergärten weitere Untergliederung nach **Halbtags-Ganztagseinrichtungen**	Unterrichtsstunden für die 10 Bildungssektoren. Anzahl der Studierenden für die Fächergruppen im Bereich Universitäten und Fachhochschulen Betreuungsstunden im Bereich Kindergärten	Durchschnittskosten	keine	**Ergebnis der Unterscheidung in Sektoren:** Ab 1996 (-2001) etwas höherer Volumenanstieg (2.6% vs. 1.1%) als ohne Untergliederung, was auf Volumenanstieg bei Sektoren zurückzuführen ist, welche höhere Durchschnittspreise verzeichnen. (höhere Anteile von Kindergärten im Sektor der Privaten Organisationen und Weiterbildung im Sektor Unternehmen) **Ergebnis Unterteilung in Universitäten:** Im Zeitraum 1992-2001 hat sich das Volumen um 7,1 Indexpunkte verringert (im Gegensatz zu 2.3% bei ungewichteter Zahl der Studierenden), was auf eine Verschiebung in Richtung kostengünstigere Studiengänge zurückzuführen ist. (Qualitätsstruktureffekt). **Unterscheidung Input vs. Outputmethode insgesamt:** Über den Zeitraum 1995-2003 weist die direkte Outputmethode einen höheren Volumenanstieg auf (4.5% vs. 3.8 %)
Finnland	**2 Sektoren** (zentral und lokal) untergliedert nach **6 Bildungsstufen** (+Erwachsenen- and sonstige Bildung). Bildungsstufen weiter untergliedert in **reguläre und Sonderschüler** Universitäten werden unterteilt in **20 Fächergruppen,** höhere Abschlüsse, Forschung; Polytechnischen Schulen nach **7 Fächergruppen**	Zahl der Schüler Für Universitäten und Polytech-Schulen **Zahl der erworbenen Kreditpunkte** Zahl der Studenten für höhere Abschlüsse Zahl der Publikationen für Forschung	Durchschnittskosten	Implizite Adjustierung für Universitäten und Ploytech. Schulen anhand von Outcome- bzw. Erfolgsmaßen (Kreditpunkte)	Für **Polytech** Zahl der Studenten oder Kreditpunkte sehr geringer Unterschied. Größere Differenz zwischen gewichteten und ungewichteten Maßen. **Gewichtung führt zu höherem Output über die Periode 2000-2005 (ca. 2% höher in 2005)** Bei Universitäten führt **Kreditpunktmessung zu signifikant höherem Output als die Zahl der Studenten (ca 4% in 2005)**
Italien	Bildungsdienstleistungen untergliedert in **4 Gruppen** (Schule, Ausbildung, Universität und Hilfsdienst-	Zahl der Schüler, Zahl der Studenten	Durchschnittskosten	Schulsystem: **Klassengröße**: Verwendung einer **Kongestions-Funktion** um die tatsächliche	Keine Angaben

	leistungen). Schule untergliedert in **4 Bildungsstufen** (pre-primary, primary, lower und upper secondary) Upper secondary weiter unterteilt in 7 Schultypen Universitäten unterteilt in **18 Fakultäten**			Schülerzahl pro Klasse in „standard“ Schülerzahlen pro Klasse zu transformieren (Min-Max Klassengrößen institutionell vorgegeben) Universitätssystem: 2 Indikatoren: a) Quotient aus der Zahl der Studenten, die Regelstudienzeit nicht überschritten haben, zu allen Studenten. b) Quotient aus Regelstudienzeit und tatsächlicher Studienzeit. Gleichgewichtung beider Indikatoren	
Schweden	Gesetzlich vorgeschriebene Bildungsstufen und höher Sekundarstufen, als auch Vorschul-Aktivitäten unterteilt nach Dienstleistungen: (1) Lehre (2) Schulmalzeiten (3) Schultransport (4) Schülerwohlfahrt (z.B. Arztbesuch) (5) Andere Dienst (6) Vorschule (7) Sprachkurse Universitärer Bereich separat erfasst	(1) Zahl der Schüler (2) Zahl der Schüler (3) Volumen-Index über deflationierte nominale Werte (4) Zahl der Schüler inklusive Zahl er Schüler im 9. Schuljahr doppelt gezählt (Studienwahl-Unterstützung nur im 9. Schuljahr) (5) Zahl der Schüler (6) Zahl der Kinder (7) Korrektur um Faktor 0.2	Kostengewichte für einzelne Quantitätsmaße (1)–(7)	(1) Adjustierung der Zahl der Schüler, die das 9. Schuljahr: (a) ohne, (b) mit Prädikatsexamen, (c) um Prozentsatz von Studenten die sich für eine Ausbildung in der Bildungsstufe upper secondary qualifiziert haben	Für gesetzlich vorgeschriebene Bildungsstufen: Qualitätsadjustierung führt zu einem marginalen Anstieg des Volumen-Index von 99,96 auf 100,2 (Ergebnis einer Beispielsrechnung)
Großbritannien	(1) Primäre, sekundäre, und „spezial“ Schulen, City Technical Colleges und Akademien (2) Vorschule (3) Studenten in den Studiengängen Gesundheit und ITT (Initial teacher training)	(1) Anwesenheit der Schüler (2) Anwesenheit der Schüler (3) Studentenzahlen	Kostengewichte für einzelne Quantitätsmaße (1)–(3)	(1) jährliche Qualitätsadj.: 0.25% (historische Verbesserungsrate in den GCSE Testresultaten seit Mitte der 90er) (3) Anteil von Studenten im letzten Jahr, die QTS erhalten werden	Qualitätsadj.: Bildungs-Output zw. 1996-2006: 5.8% (mit Qualitätsadj., multiplikative Berechnungsweise) 3.8% (ohne Qualitätsadj., additive Berechnungsweise) [Nicht in offi. Statistiken verwendet: weitere Qualitätsadjustierungen, additive Berechnungsweise: 27.1% (Testlauf)]

Quellen: Destatis (2004, 2005), Statistics Finland (2007), ISTAT (2007), UK National Statistics (2007), Statistics Sweden (2007).

5.3 Simulationen von Outputindizes des Bildungssektors

Die im Folgenden dargestellten Simulationen für unterschiedliche Berechnungen von Outputindizes erfolgten aufgrund der bereits angesprochenen Tatsache, dass sich bisher keine internationalen Standards herausbilden konnten und unterschiedlich propagierte Methoden häufig zu voneinander abweichenden Ergebnissen führen. Im Rahmen der Simulationen sollen deshalb die unterschiedlichen Methoden und ihre quantitativen Auswirkungen auf den Outputindex des Bildungssektors näher untersucht werden.

Hier fokussieren wir uns auf Deutschland. Simulationsergebnisse für andere Länder finden sich im Anhang dieser Studie, wobei die Wahl der Länder von der Verfügbarkeit des entsprechenden Datenmaterials abhängig gemacht wurde. Die verwendeten Daten stammen hauptsächlich aus Datenbanken der UNESCO, da öffentlich verfügbare Datenquellen für disaggregierte Analysen schwer verfügbar sind und mit ihnen keine internationale Vergleichbarkeit gewährleistet werden kann. Gerade aber die Vergleichbarkeit über mehrer Länder hinweg unterstützt die Unabhängigkeit und somit die Robustheit der untersuchten Methoden.

Die Simulationen beginnen mit der Illustration des Verlaufs eines Volumenindex für den Bildungssektor über den Zeitraum 1999–2005, der auf das Basis- bzw. Referenzjahr 2001 indiziert wurde und der reale Veränderungen in Bezug auf das Referenzjahr anzeigt. Beginnend erfolgt eine Darstellung für einen Volumenindex der ausschließlich auf der ungewichteten Anzahl eingeschriebener Schüler und Studenten innerhalb einzelner Ausbildungsstufen basiert. Dabei werden die Ausbildungsstufen nach vier ISCED-Klassifikation[25] untergliedert. Diese belaufen sich auf Kindergarten (pre-primär), Grundschule (Primärstufe), Sekundarstufen (Haupt-, Real-, Fachober- und Berufsoberschule, sowie Gymnasium) und Tertiärstufen (Fachhochschule und Universitäten). Die darauf folgenden Indizes sind Modifikationen dieses ersten Indikators. Der nächste Index repräsentiert einen kostengewichteten Outputindex, wobei die einzelnen Bildungsstufen nun unter Verwendung von Kostengewichten aggregiert werden. Es folgen Indizes mit Outcome basierten Quantitätsindikatoren des tertiären Sektors, sowie mit entsprechenden Qualitätsadjustierungen für nicht-tertiäre Bereiche. Zur Vergleichbarkeit mit offiziellen Statistiken wird ebenfalls der Volumenindex für den Bildungssektor des Statistischen Bundesamtes abgetragen.

[25] Der ISCED (International Standard Classification of Education) Standard wurde von der UNESCO zur Klassifikation von Schultypen und -systemen entwickelt. Dabei wird zwischen mehreren Ausbildungsebenen unterschieden.

Ungewichtete und kostengewichtete Outputindizes

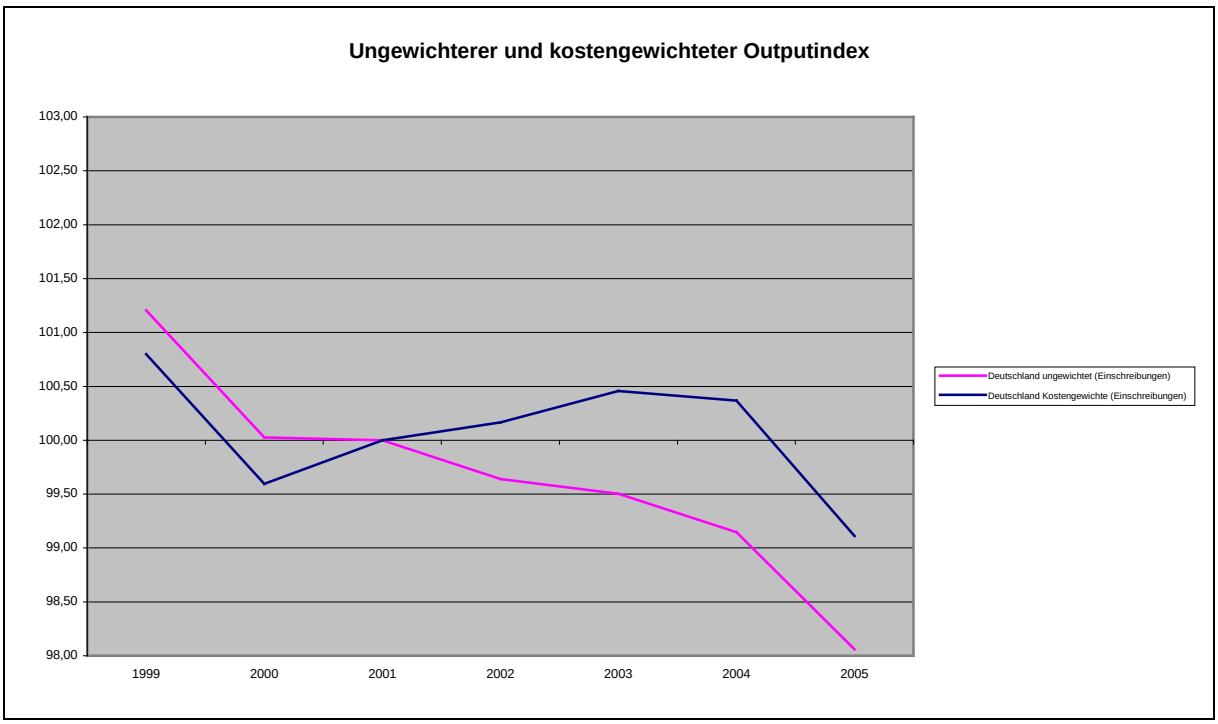
Bei der Konstruktion des ungewichteten Outputindex greifen wir auf die Gesamtzahl aller Schüler und eingeschriebenen Studenten, unabhängig ihrer Zugehörigkeit zu einzelnen ISCED-Ausbildungsstufen, zurück (siehe Abbildung 5.2). Der sich daraus ergebende Mengenindex[26] zeigt im Hinblick auf das Basisjahr 2001 einen stetigen Rückgang über den gesamten Zeitraum. Im Jahr 2005 weist er gegenüber 2001 einen um etwas 2 Indexpunkte geringeren Wert auf (100 gegenüber 98,06).

Als Alternative zu der ungewichteten Berechung kann die Anzahl der Schüler und Studenten in den einzelnen Bildungsstufen aber auch gewichtet mit ihren jeweiligen Kostenanteilen aggregierte werden. Die Wahl von Kostengewichten lehnt sich an die bisherige Vorgehensweise zur Aggregation in Deutschland und in vergleichbaren Industrienationen an (vgl. 5.2 Methodenüberblick). Trotz der bereits angesprochenen Probleme bei der Verwendung von Kostengewichten sind diese oft die einzig verfügbare Alternative zur Approximation von Preisgewichten. Durch die explizite Gewichtung weist dieser Volumenindex nun einen Indexwert von 99,11 Indexpunkten in 2005, d.h. einen um 0,89 Indexpunkte niedrigeren Wert als im Basisjahr 2001 auf.

Die Methode der Kostengewichtung führt dazu, dass der Volumenindex im Jahr 2005 ein um etwa 1,0 Indexpunkte höheres Outputwachstum generiert als der einfache ungewichtete Mengenindex (99,11 vs. 98,06). Der Grund für den höheren Outputwert liegt in dem deutlich höheren Gewicht, das den Schülern im sekundären Bildungsbereich zukommt. Dabei ist das Gewicht der Sekundarstufen so hoch wie das Gewicht der pre-primären, primären und tertiären Bildungsstufen zusammen. Die Veränderungen in den Schülerzahlen sind dabei nicht ausschlaggebend, da sie sich nicht ungewöhnlich stark verändert haben. In den selektierten Ländern zeigen die Simulationen nur für Schweden und Großbritannien eine Erhöhung des Bildungsoutputs in 2005 unter Verwendung der Kostengewichte gegenüber den umgewichteten Mengenindizes. Für Italien und Frankreich dagegen weisen die Indizes mit Kostengewichten jeweils eine niedrigere Outputentwicklung in 2005 aus, wobei sich nur marginale Differenzen in den Wachstumsraten zwischen beiden Indexmethoden konstatieren lassen (siehe Anhang).

[26] Die in dieser Studie verwendeten Indizes basieren auf dem Prinzip der Log-Change Indizes. Bei der Implementierung von Kostengewichten wird der Thörnqvist-Index implementiert.

Abb.: 5.2: Output-Entwicklung im Bildungssektor, Deutschland (Volumenindex 2001=100)



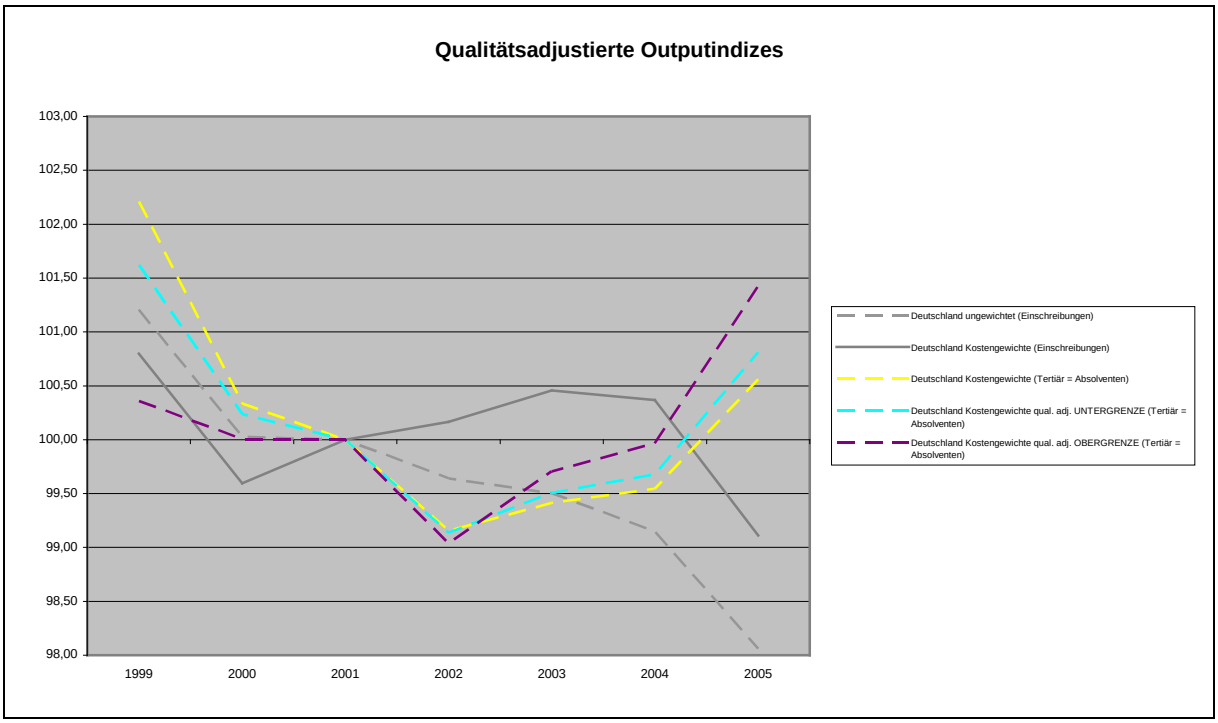
Quellen: UNESCO (2008), eigene Berechnungen.

Alternativ zu den Kostengewichten ließe sich auch wie im oben beschriebenen Beispiel Schottlands eine Gewichtung mittels Outcome basierter Indikatoren durchführen. Dabei werden für die Zukunft zu erwartende Löhne bzw. Lohndifferenziale nach Abschluss einer bestimmten Ausbildungsstufe als Gewichte herangezogen. Zur Bestimmung der Mengenkomponente verlangt dieser Ansatz Angaben über die von den Schülern abgeschlossenen Ausbildungsstufen. Während Lohndifferenziale für einzelne Bildungsstufen über die OECD erhältlich sind, ist die Beschaffung von Daten hinsichtlich abgeschlossener Ausbildungsstufen von Schülern weitaus schwieriger. Die zum Zeitpunkt der Studie vorhandenen Daten bezogen sich ausschließlich auf die Anzahl der eingeschriebenen Schüler innerhalb einzelner Ausbildungsstufen. Eine Verwendung dieser Daten ist jedoch nicht ratsam, da man dabei unterstellen würde, dass alle Schüler, die in einer bestimmten Ausbildungsstufe registriert sind, diese Stufe letztlich bestehen, und zudem, dass sie nur diese eine Ausbildungsstufe, in der sie sich gerade befinden, absolvieren. Aufgrund der fehlenden Daten für einer korrekte Umsetzung dieser Methode und der mit den vorhandenen Daten verbundenen restriktiven Annahmen wird von dieser Art der Qualitätsadjustierung abgesehen.

Explizite Qualitätsadjustierung in den Outputindizes

Die Einführung von Kostengewichten hat bereits erste Veränderungen in den Outputindizes aufgezeigt. Um die Veränderung der Outputindizes unter Qualitätsadjustierungen zu verdeutlichen, werden im Folgenden zunächst anstelle der eingeschriebenen Studentenzahlen die Anzahl der Absolventen zur Messung des Outputs im tertiären Bildungssektor herangezogen. Die Intention dieser Anpassung liegt darin, anstelle des Aktivitätsindikators den Output direkt über die Verwendung eines Outcome-Indikators zu messen und somit die Qualität einer Leistung direkt mit der Wahl des Quantitätsmaßes zu erfassen. Die Implementierung des Outcome-Indikators reflektiert zudem eine exaktere Messung des universitären Outputs, da er sich ausschließlich auf die abgeschlossenen Studiengänge bezieht und somit Studienabbrecher außer Acht lässt. Die Verwendung von Outcome basierten Quantitätsmaßen birgt jedoch auch Schwierigkeiten der Zurechenbarkeit, die wir bereits oben erörtert haben.

Abb.: 5.3: Output-Entwicklung im Bildungssektor, Deutschland (Volumen-Index 2001=100)



Quellen: UNESCO (2008), eigene Berechnungen.

Wie in Abb. 5.3 deutlich wird, zeigt der Übergang von Aktivitäts- zu Outcome basierten Mengenindikatoren bereits einen signifikanten Einfluss auf die Entwicklung des Outputindex. Die Qualitätsadjustierung der tertiären Bildungsstufe führt dazu, dass der Indexwert in 2005 um ca. 0,5 Indexpunkte höher lag als im Basisjahr. Verglichen mit den kosten- bzw. ungewichteten Outputindizes konnte sich der Output dementsprechend um 1,5 bis 2,5 Indexpunkte

(100,56 zu 99,11 bzw. 98,06) erhöhen. Die Umkehr in der Trendentwicklung ist dabei auf die starke Zunahme akademischer Abschlüsse im Gegensatz zu rückläufigen Studentenzahlen zurückzuführen. Die Umstellung von Studenten- auf Absolventenzahlen führt in allen Ländern zu einer Erhöhung des Outputwerts (siehe Anhang). Die Simulationen zeigen bereits die Sensitivität des Outputmaßes aufgrund der Umstellung auf Outcome basierte Quantitätsindikatoren.

Neben der Qualitätsadjustierung des tertiären Sektors, sollen im Folgenden noch Qualitätsveränderungen im primären und sekundären Bildungssektor berücksichtigt werden. Für diese Bildungssektoren soll auf eine Qualitätsadjustierung der Schülerzahlen mittels des Lehrer-Schüler-Verhältnisses zurückgegriffen werden. Die Einführung einer derartigen Qualitätskomponente zum Aktivitätsindex wird beispielsweise im *Atkinson Report* vorgeschlagen. Die Schwierigkeit besteht darin, das Lehrer-Schüler-Verhältnis in einen Qualitätsfaktor umzurechnen, der dem Output multiplikativ zugerechnet werden kann. Hierzu müssen Elastizitäten berechnet werden, die beschreiben, in welchem prozentualen Umfang eine 1% Erhöhung (Verringerung) des Lehrer-Schüler-Verhältnisses den Output steigert (senkt). In unserer Simulation verwenden wir die Ergebnisse von Studien aus den Vereinigten Staaten[27]. Basierend auf unterschiedlichen Annahmen über den Effekt der Klassengröße auf die Ergebnisse von Leistungstest sowie von Leistungstests auf den Schuloutput ergeben sich aus diesen Studien eine Ober- und Untergrenze für diese Elastizität.[28] Die Untergrenze der Elastizität entspricht 0.1%, d.h., dass eine einprozentige Reduktion der Klassengröße einer Erhöhung des Schuloutputs um 0,1 Prozent entspricht. Für die Obergrenze wird eine Elastizität von 0.33% angenommen. Die Verwendung dieser beiden alternativen Elastizitäten resultiert in einer Ober- bzw. Untergrenze für den Outputindex.

Abb. 5.3 zeigt die durch Qualitätsadjustierung des Schuloutputs induzierten Ober- und Untergrenze des Outputs. Dabei liefert die Untergrenze in 2005 einen Indexwert, der um 0,81 Indexpunkte über dem des Basisjahres liegt (100,81 zu 100), während die Obergrenze einen um 1,42 Indexpunkte höheren Output konstatieren lässt (101,42 zu 100). Durch die Qualitätsadjustierung im Schulsektor liegen die beiden Werten für die Unter- und Obergrenze mit 0,25 bzw. 0,86 Indexpunkten über dem vorherigen, ausschließlich für den tertiären Bereich qualitätsadjustierten Outputindex (100,81 bzw. 101,42 zu 100,56 in 2005). Die Erhöhung des Out-

[27] Christian (2006).

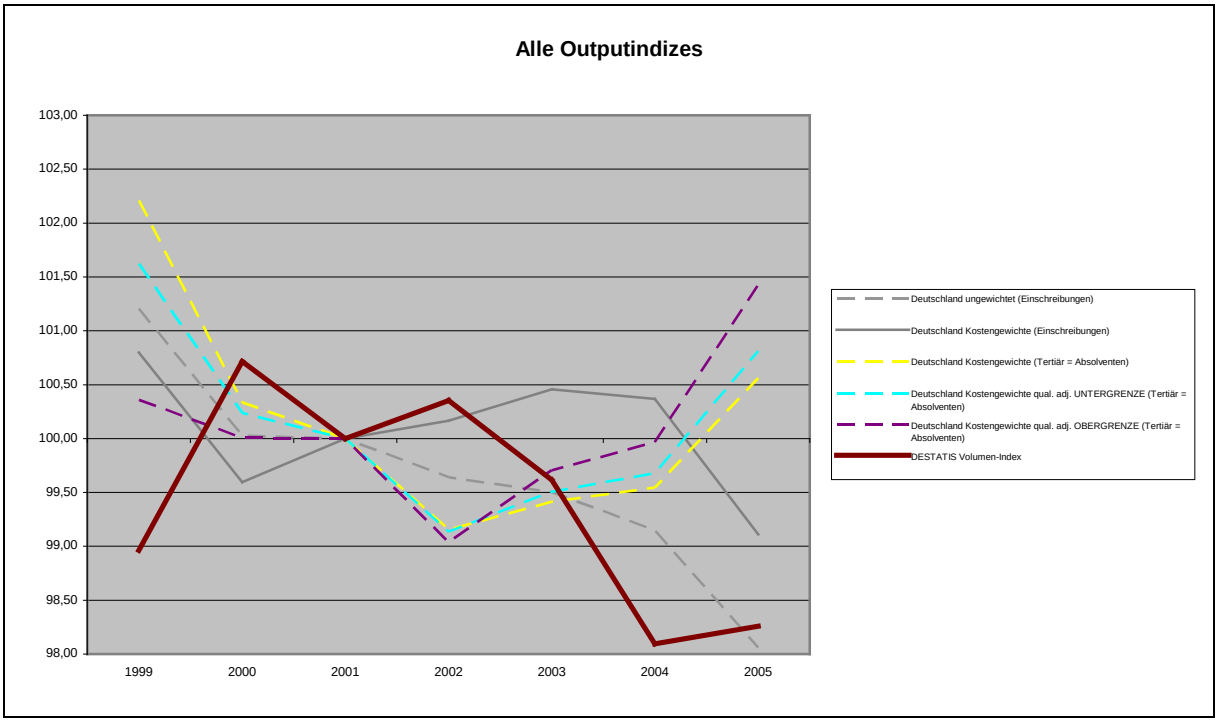
puts durch die zusätzliche Einführung einer Qualitätsadjustierung des Schuloutputs ist auf die unterstellten Outputelastizitäten einer Reduktion in der Klassengröße zurückzuführen. Dabei wirken die Elastizitäten wie Skalierungseffekte, wobei höhere Elastizitäten einen größeren positiven Einfluss auf den Schuloutput und damit den gesamten Bildungsoutput ausüben. Für Frankreich, Italien und Großbritannien liefert die zusätzliche Qualitätsadjustierung des Schuloutput mittels Bildung von Unter- und Obergrenzen nur marginale Veränderung. Dagegen zeigen sich in Schweden deutlich höhere Outputentwicklungen unter der Verwendung dieser Methode (siehe Anhang).

Zusätzlich zu den in dieser Studie simulierten Output-Indizes wird der vom statistischen Bundesamt veröffentlichte reale Volumenindex abgetragen. Wie in Abb. 4.4 deutlich sichtbar wird, weist der offizielle Volumenindex gerade in den letzten Jahren eine deutlich schlechtere Outputentwicklung des Bildungssektors auf, als die von uns berechneten Outputindizes. In 2005 liegt sein Indexwert bei 98,26. Auch bleibt zu konstatieren, dass der rückläufige Verlauf des offiziellen Outputindex bereits in früheren Jahren eingesetzt hat und eine Umkehr in der Trendentwicklung nicht erkennbar ist. Im Vergleich zu den qualitätsadjustierten Outputindizes liegt der Indexwert des offiziellen Volumenindex in 2005 mit 98,26 zu 100,56 (Qualitätsadjustierung ausschließlich im tertiärer Bereich), 100,81(Qualitätsadjustierung zusätzlich des Schuloutputs, Untergrenze) bzw. 101,42 (Qualitätsadjustierung zusätzlich des Schuloutputs, Obergrenze) deutlich niedriger.

Abschließend ist bei einer derartigen Gegenüberstellung aber darauf hinzuweisen, dass die Berechnungsweisen der zu Simulationszwecken konstruierten und des offiziell veröffentlichen Outputindizes signifikant voneinander abweichen können. Da die Berechnung des offiziellen Volumenindex auf einer deutlich höheren Disaggregationstiefe basiert als die von uns in erster Linie zu Illustrationszwecken verwendeten ISCED-Klassifikationen, ist davon auszugehen, dass der offizielle Index gerade im Hinblick auf die abgebildeten Qualitätsstruktureffekte vergleichsweise exakt berechnet ist. Zudem muss berücksichtigt werden, dass die Verwendung von UNESCO-Daten nicht zwangsläufig mit den dem Statistischen Bundesamt zur Verfügung stehenden Daten übereinstimmen, wodurch die Resultate voneinander abweichen können. Die konträren Verläufe, insbesondere vor dem Basisjahr, zeigen die Schwierigkeit, die mit der Erzeugung konsistenter Berechnungen bei unterschiedlicher Ausganglage des Datenmaterials verbunden sind. Korrelationsanalysen zwischen dem offiziellen Volumenindex und dem von uns konstruierten kostengewichteten Outputindex zeigen jedoch, dass die

beiden Zeitreihen bereits ab und nach dem Basisjahr recht ähnliche Verläufe aufweisen. Ab dem Basisjahr beläuft sich der Korrelationskoeffizient auf etwa 0,4, mit zunehmender Tendenz. Dessen ungeachtet verdeutlicht die Gegenüberstellung alternativer Outputberechnungsmethoden, nicht nur gegenüber offiziellen Statistiken, sondern auch im Rahmen heuristisch deduzierter Indexkonstruktionen, die Auswirkungen bereits geringfügiger Qualitätsadjustierungen auf die Entwicklung realer Outputgrößen.

Abb.: 5.4: Output-Entwicklung im Bildungssektor, Deutschland (Volumen-Index 2001=100)



Quellen: UNESCO (2008), Destatis (2008), eigene Berechnungen.

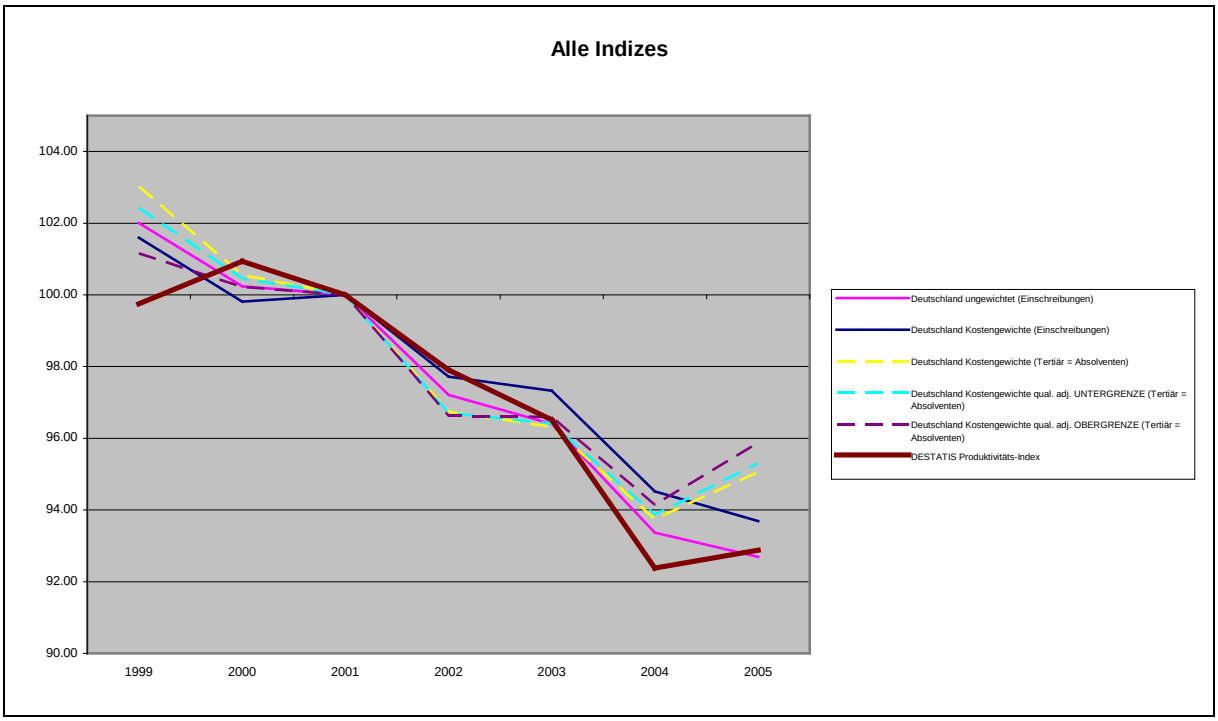
5.4 Simulationen zur Produktivitätsentwicklung

Zur Durchführung von Produktivitätsanalysen bedarf es neben der Konstruktion geeigneter Outputindizes, ebenso der Berücksichtigung der Inputseite. Wie in den vorherigen Kapiteln angesprochen können Produktivitätskennzahlen auf Basis so genannter Ein-Faktor- bzw. Multi-Faktor-Produktivitätskennziffern abgeleitet werden. Aufgrund der geringeren Datenmenge wird im Folgenden die Berechung eines Ein-Faktor-Produktivitätsmaßes, d.h. der Arbeitsproduktivität, durchgeführt. Dazu wird der reale Output (Volumen) ins Verhältnis gesetzt zur Anzahl der gearbeiteten Stunden[29] im Bildungssektor. Die graphische Illustration der Produk-

[29] Die gearbeiteten Stunden beziehen sich dabei auf die Arbeitnehmer, d. h. nur die vom Staat direkt angestellten Erwerbspersonen.

tivitätsindizes zum Basisjahr 2001 (=100) ist in Abb. 4.5 dargestellt und bezieht sich auf die zuvor berechneten Outputindizes des Bildungssektors, einschließlich der von statistischen Bundesamt veröffentlichen Volumengrößen.

Abb.: 5.5: Arbeitsproduktivität im Bildungssektor, Deutschland (Index 2001=100)



Quellen: UNESCO (2008), Destatis (2008), eigenen Berechnungen.

Während sich die unterschiedlichen Produktivitätsindizes um das Basisjahr herum nicht wesentlich voneinander unterscheiden und auch die Verläufe grundsätzlich sehr ähnlich sind, zeichnen sich ab 2003 dennoch divergente Entwicklungstendenzen ab. Es fällt auf, dass die unter Berücksichtigung von Qualitätsadjustierungen konstruierten Produktivitätsindizes deutlich weniger negative Verläufe im Indexwert aufweisen. Während nach offiziellen Statistiken die Produktivität im Bildungssektor im Vergleich zum Basisjahr auf einen Wert von 92,88 Indexpunkten in 2005 zurückfällt, liegt dieser Wert nach der Implementierung Outcome basierter Quantitätsindikatoren bei 95,06. Die zusätzliche Qualitätsadjustierung des Schulsektors lässt für die Unter- und Obergrenzen einen Indexwert von 95,29 bzw. 95,87 in 2005 konstatieren. Maßgeblich für die höhere Bandbreite in der sich die Produktivitätsindizes im Vergleich zu den Outputindizes befinden, ist die starke Zunahme der gearbeiteten Stunden (Inputseite), zu denen alle Indizes ins Verhältnis gesetzt werden.

6. Schlussfolgerung

Die Messung des Beitrags des öffentlichen Sektors zur Wertschöpfung ist, vor allem bezüglich des Verlaufs, eine wesentliche Voraussetzung für die Beurteilung gesamtwirtschaftlicher Zusammenhänge und für die Bewertung der Lohn- und Personalpolitik im öffentlichen Sektor. Traditionell wird ausschließlich auf die Inputseite abgestellt. Dies ist problematisch, da öffentliche Ausgaben stets von der Kostenseite her betrachtet werden, selbst wenn die Ausgaben für das Gemeinwesen sehr förderlich sind. So besteht die Gefahr, dass Kürzungen beim Mitteleinsatz einseitig auch dort als Einsparungen gewertet werden, wo sie vielleicht das Leistungsangebot schwer beeinträchtigen. Zudem fehlt eine Maß für die Produktivitätsentwicklung.

Bislang ist die amtliche Statistik von geringen Steigerungen in der Produktivität des öffentlichen Sektors ausgegangen, die deutlich unter den Durchschnittswerten in anderen Branchen liegen. Dies entspringt der traditionellen Sicht, wonach im Dienstleistungsbereich generell nur geringe Produktivitätszuwächse zu verzeichnen sind. Angesichts des erkennbaren Modernisierungsfortschritts im öffentlichen Sektor ist diese Sicht zu hinterfragen. Der verstärkte EDV Einsatz, eine veränderte Zusammenarbeit mit dem privaten Sektor und neue Organisationsstrukturen legen viel mehr Produktivitätszuwächse nahe.

Ziel der vorgelegten Untersuchung ist es vor diesem Hintergrund, die Problematik der Messung der Leistung des öffentlichen Sektors zu untersuchen und Ansätze für eine verbesserte Methode zur Bestimmung der Leistungsfähigkeit des öffentlichen Sektors zu diskutieren. Letzteres bezieht sich vor allem auf die so genannte Outputmethode. Sie greift auf Mengenindikatoren für Dienstleistungen zurück und aggregiert dann den so gemessenen Output in den einzelnen Bereichen durch geeignet gewählte Gewichte.

Die im letzten Kapitel angestellten Methodenvergleiche haben international eine Abkehr von traditionellen Inputverfahren hin zu einer solchen direkten Messung des Outputs gezeigt. Der Vergleich zeigt jedoch auch, dass teilweise große Unterschiede in der konkreten Umsetzung der Outputmessung insbesondere bei der Berücksichtigung von Qualitätsveränderungen bestehen. Dies ist zum einen auf Datenrestriktionen und zum anderen auf fehlende internationale Standards zurückzuführen. Die Simulationen haben darüber hinaus gezeigt, dass verschiedene Ansätze der direkten Outputmessung zu signifikant unterschiedlichen realen Outputentwicklungen im Bildungssektor führen können. Unter dem Vorbehalt, dass die in dieser Studie an-

gestellten Simulationen aufgrund unterschiedlicher Disaggregationstiefen und Datenbestände keine exakten Vergleiche mit offiziellen Statistiken zulassen, lässt sich dennoch feststellen, dass die explizite Berücksichtigung von Qualitätsveränderungen generell zu einer Erhöhung des gemessenen Outputs führt. Dies wird auch durch die bisher durchgeführten Studien der statistischen Ämter der hier untersuchten Länder bestätigt.

Da in Deutschland die Qualität des Bildungssektors bisher nur implizit über so genannte Qualitätsstruktureffekte gemessen wird, legen unsere Ergebnisse den Schluss nahe, dass die offiziell ermittelten Werte für Output und Produktivität des Bildungssektors eine Untergrenze bilden. Explizite Qualitätsadjustierungen mittels geeigneter Outcome-Indikatoren könnten zu weitaus positiveren Schlussfolgerungen über die tatsächliche Produktivitätsentwicklung im öffentlichen Sektor der letzten Jahre führen, als sie bisher angestellt wurden.

Um verlässlichere Aussagen über Output und Produktivitätsentwicklungen im öffentlichen Sektor treffen zu können, ist es unverzichtbar, auf internationaler Ebene gewisse „best practices" zu identifizieren. Darauf aufbauend müssten die statistischen Ämter der einzelnen Länder verpflichtet werden, die benötigten Daten zu erheben. In diesem Zusammenhang ist die Arbeit der OECD im Rahmen des Projekts *Measuring Education and Health Volume Output* hervorzuheben, in dem versucht wird, genau diese „best practices" zu identifizieren.[30]

[30] Informationen und erste Veröffentlichungen zu diesem Projekt sind verfügbar im Internet unter http://www.oecd.org/document/47/0,3343,en_2825_495684_37733615_1_1_1_1,00.html, [Stand 29.06.2008].

Referenzen

Atkinson, T. (2005), „Atkinson Review: Final Report: Measurement of Government Output and Productivity for the National Accounts", Palgrave McMillian

Baumol, W.J.(1967), "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis", American Economic Review, 57, 3 (1967), S. 415-426.

Boskin Report (1996), "Advisory Commission To Study The Consumer Price Index – Towards a More Accurate Measure of the Cost of Living", Final Report to the Committee on Finance of the U.S. Senate.

Braakmann, A., N. Hartmann, N. Räth und W. Strohm (2005), „Revision der Volkswirtschaftlichen Gesamtrechnung 2005 für den Zeitraum 1991 bis 2004", Wirtschaft und Statistik, 5/2005, Statistisches Bundesamt, Wiesbaden.

Christian, M. (2006), "Measuring the Education Function of Government in the United States", verfügbar im Internet unter http://www.oecd.org/dataoecd/1/62/37460198.pdf [Stand: 04.07.2008].

Collesi, D., Guerrucci, D., Versace, D. und S. Zannoni (2007), „The Use of Class Size and the Italian Method", vorbereitet für das OECD Handbook "Measuring Education and Health Volume Output", verfügbar im Internet unter http://www.oecd.org/dataoecd/51/8/38685528.pdf [Stand: 14.07.2008].

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", Journal of Econometrics, 4, S. 115–145.

Diewert, W. E. (2008), "The Measurement of Nonmarket Sector Outputs and Inputs Using Cost Weights", Discussion Paper 08-03, Department of Economics, University of British Columbia.

Diewert, W.E. (1978), "Superlative index numbers and consistency in aggregation", Econometrica, 46, S. 883–900.

Diewert, W. E. (1987), "Index Numbers", in John Eatwell, Murray Milgate und Peter Newman, *The New Palgrave: A Dictionary of Economics*, Vol. 2, S. 767–780.

Divisia, F. (1926), "L'indice monetaire et la theorie de la monnaie", Societe anonyme du Recueil Sirey, Paris.

Eichhorn, W. und J. Voeller (1983), "Axiomatic Foundations of Price Indices and Purchasing Power Parities", in Diewert W.E. und C. Montmarquette, Hrsg., Price Level Measurement, Minister of Supply and Services, Canada, S. 411–450.

Eurostat (2001), "Handbook of Price and Volume Measures of National Accounts".

Fisher, I. (1922), "The Making of Index Numbers", Boston, Houghton Mifflin.

Heikkinen, J. und S. Hautakangas (2007), “The Use of Credits and Fields of Education in Tertiary Education - the Finnish Method“, vorbereitet für das OECD Handbook “Measuring Education and Health Volume Output”, verfügbar im Internet unter http://www.oecd.org/dataoecd/57/37/38606576.pdf [Stand: 14.07.2008].

Konüs, A.A. (1924), “The Problem of the True Index of the Cost of Living”, Econometrica, Nr. 7, (1939), S. 10–29.

Laspeyres, E. (1871), “Die Berechnung einer mittleren Warenpreissteigerung”, Jahrbücher für Nationalökonomie und Statistik, 16, S. 296–315.

Magnusson, B. (2007), „An Example of Quality Adjustment in Education from Sweden”, vorbereitet für das OECD Handbook “Measuring Education and Health Volume Output”, verfügbar im Internet unter http://www.oecd.org/dataoecd/28/40/38699029.pdf [Stand: 14.07.2008].

Office for National Statistics (ONS) (2007), “Public Service Productivity: Education”, verfügbar im Internet unter http://www.statistics.gov.uk/articles/nojournal/Education_productivity_2007_main.pdf [Stand: 14.07.2008].

OECD (2001), “Measuring Productivity: OECD Manual: Measurement of Aggregate and Industry-Level Productivity Growth”, OECD, Paris.

OECD (i.E.), “Handbook on Measuring Output in Health and Education”, OECD, Paris.

Paasche, H. (1874), “Über die Preisentwicklung der letzten Jahre nach den Hamburger Börsennotierungen“, Jahrbücher für Nationalökonomie und Statistik, 23, S. 168–178.

Samuelson, P. (1938), “A Note on the Pure Theory of Consumers' Behaviour”, Economica, Nr. 5, S. 61–71.

Thörnqvist, L. (1936), “The Bank of Finland’s Consumption Price Index”, Bank of Finland Monthly Bulletin 10, S. 1–8.

Triplett, J. (2006), “Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products”, Statistical Analysis of Science, Technology and Industry OECD, Paris.

Triplett, J. und Bosworth, B. (2003), Productivity measurement issues in services industries: "Baumol's Disease" has been cured, Economic Policy Review - Federal Reserve Bank of New York, Sep 2003

Anhang

A1. Ober- und Untergrenzen des theoretischen Preisindex

Nach Gleichung (6) wird unterstellt, dass die Lösung zum Kostenminimierungsverhalten der Konsumente über $\sum_{i=1}^{n} p_i q_i$ beobachtet werden kann, wobei q_i die optimalen Mengen q_i^* repräsentieren. Dies lässt sich dann für die Perioden t =0 und t=1 wie folgt verwenden:

$$C^0\left[p^0, F\left(q^{0*}\right)\right] = \sum_{i=1}^{n} p_i^0 q_i^{0*} \qquad \text{(A1)},$$

$$C^1\left[p^1, F\left(q^{1*}\right)\right] = \sum_{i=1}^{n} p_i^1 q_i^{1*} \qquad \text{(A2)}.$$

Die optimalen Lösungen für diese beiden spezifizierten Kostenminimierungsprobleme sind somit in der Realität bobachtbar. Nun existieren prinzipiell zwei unterschiedliche Herangehensweisen um einen Preisindex nach den Ausführungen unter Abschnitt 1.5.1 für Gleichung (7) zu bestimmen. Einmal lässt sich der Index in Form der Referenzperiode t=0 definieren:

$$P^0\left(p^0, p^1, F\left(q^{0*}\right)\right) = \frac{C^0\left[p^1, F\left(q^{0*}\right)\right]}{C^0\left[p^0, F\left(q^{0*}\right)\right]} \qquad \text{(A3)},$$

wobei der Nenner nach (A2) eine beobachtbare Größe darstellt. Gleichung (A3) lässt sich dann formulieren als

$$P^0\left(p^0, p^1, F\left(q^{0*}\right)\right) = \frac{C^0\left[p^1, F\left(q^{0*}\right)\right]}{\sum_{i=1}^{n} p_i^0 q_i^{0*}} \qquad \text{(A4)}.$$

Während der Nenner beobachtet werden kann, ist der Zähler ein theoretisches Konstrukt, das in der Realität so nicht vorkommt. Dies liegt in der Theorie des Konsumentenverhaltens und der danach definierten Annahme begründet, dass sich die Mengen an die Preise anpassen (Preisnehmer). Würden sich nämlich der Konsument in t=0 den Preisen p^1 gegenübersehen, so würde er die Mengen q^0 in t=0 adjustieren, um dadurch eine neue kostenminimierende Lösung q^{0*} zu erreichen. Da aber q^{0*} bereits die optimale Lösung in t=0 für $C^0\left[p^0, F\left(q^{0*}\right)\right]$ gegeben p^0 darstellt und auch nur q^{0*} für t=0 beobachtet werden kann, lässt sich demnach folgern, dass durch die Verwendung der beobachteten Mengen q^0 gilt:

$$C^0\left[p^1, F\left(q^{0*}\right)\right] \leq \sum_{i=1}^{n} p_i^1 q_i^0 \qquad \text{(A5)}.$$

Die Bildung eines Preisindizes unter Fixierung der Mengen q^{0*} (zu den jeweiligen Preisen p^0) bewertet zu den Preisen p^1, liefert folglich keine exakte Umsetzung des theoretischen als op-

timal postulierten Preisindizes unter (A4). Hier werden bereits die Grenzen aufzeigt, die sich bei der Konstruktion von ökonomisch motivierten Preisindizes unter Verwendung der in der Realität beobachteten Daten ergeben. Angewendet auf Gleichung (A4) ergibt sich

$$P^0\left(p^0, p^1, F\left(q^{0*}\right)\right) \leq \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^{0*}} = P_L \qquad \text{(A6)},$$

wonach der beobachtete Laspeyres-Preisindex die Obergrenze für den theoretischen Preisindex $P^0\left(p^0, p^1, F\left(q^{0*}\right)\right)$ darstellt.

Gleichung (7) lässt sich alternativ aber auch in folgender Form spezifizieren:

$$P^1\left(p^0, p^1, F\left(q^{1*}\right)\right) = \frac{C^0\left[p^1, F\left(q^{1*}\right)\right]}{C^0\left[p^0, F\left(q^{1*}\right)\right]} \qquad \text{(A7)}$$

Es gilt analog:

$$P^1\left(p^0, p^1, F\left(q^{1*}\right)\right) = \frac{\sum_{i=1}^{n} p_i^1 q_i^{1*}}{C^1\left[p^0, F\left(q^{1*}\right)\right]} \qquad \text{(A8)},$$

sodass in diesem Fall der Zähler beobachtbar ist, während der Nenner ein theoretisches Konstrukt darstellt, das in der Realität so nicht beobachtet werden kann. Es lässt sich wiederum folgern, dass gilt:

$$C^1\left[p^1, F\left(q^{1*}\right)\right] \leq \sum_{i=1}^{n} p_i^0 q_i^1 \qquad \text{(A9)}.$$

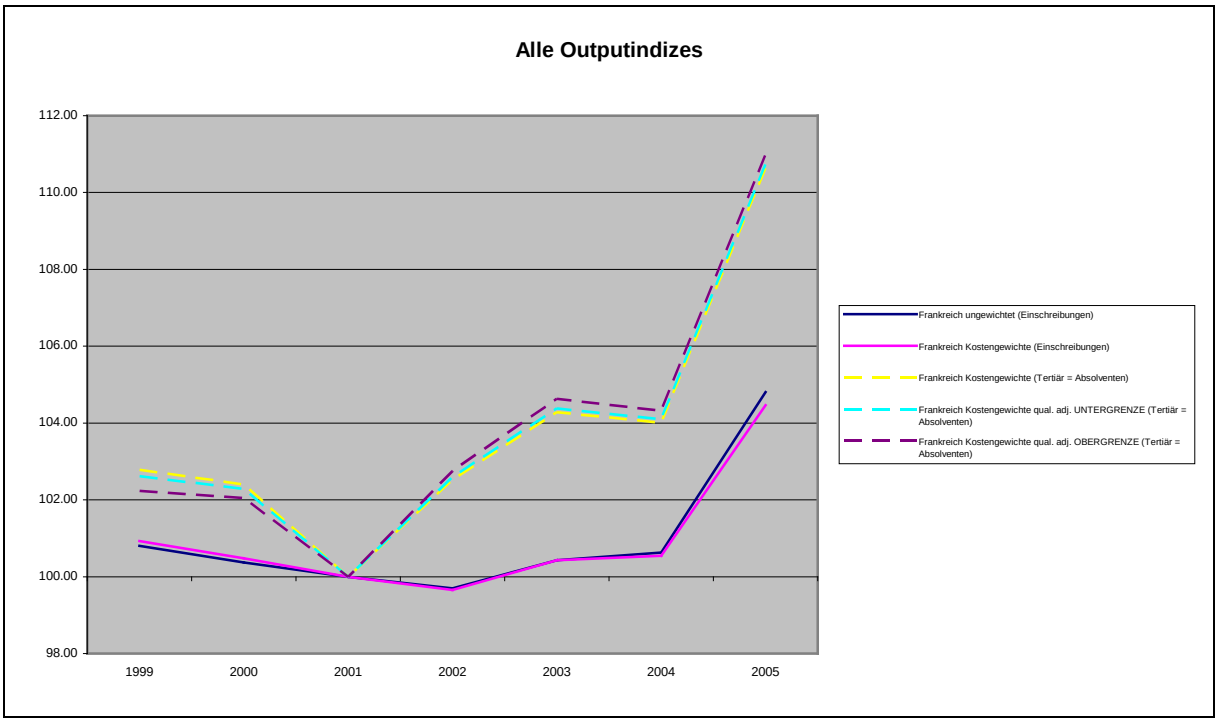
Angewendet auf Gleichung (A8) ergibt sich dann abschließend

$$P^1\left(p^0, p^1, F\left(q^{1*}\right)\right) \geq \frac{\sum_{i=1}^{n} p_i^1 q_i^{1*}}{\sum_{i=1}^{n} p_i^0 q_i^1} = P_P \qquad \text{(A10)},$$

wonach der beobachtete Paasche-Preisindex die Untergrenze für den theoretischen Index $P^1\left(p^0, p^1, F\left(q^{1*}\right)\right)$ liefert.

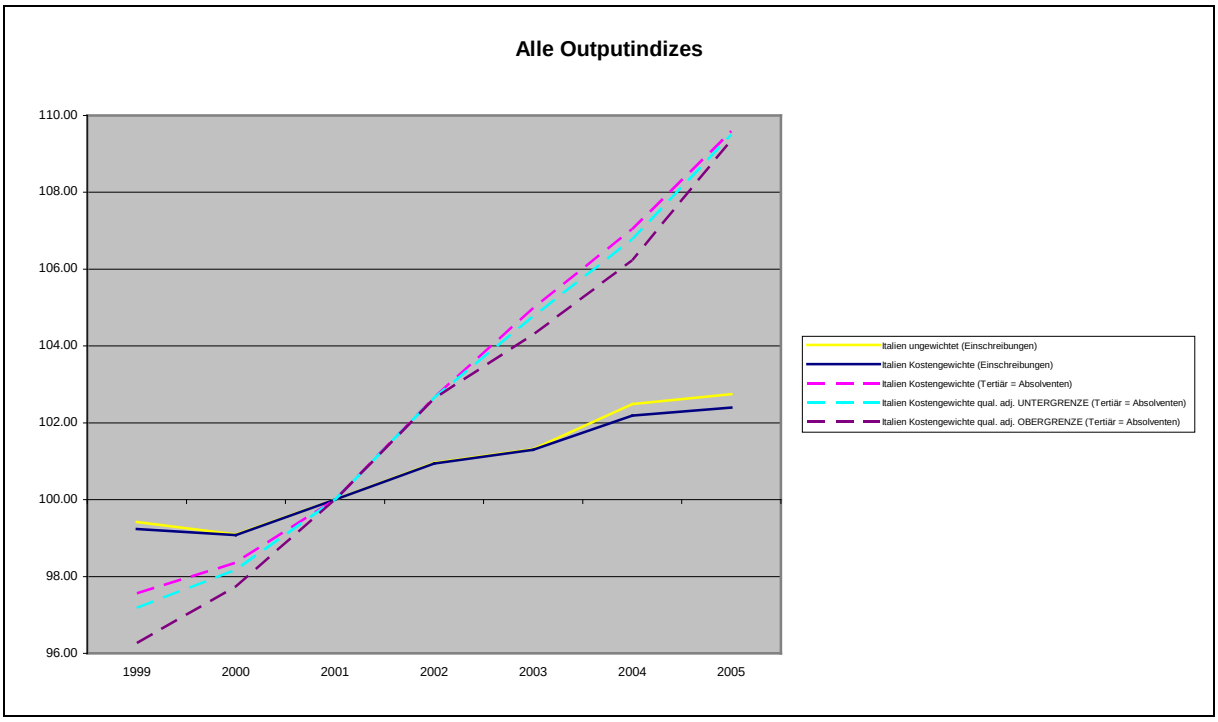
A2. Internationale Output-Indizes

Abb. A2.1: Output-Entwicklung im Bildungssektor, Frankreich (Volumen-Index 2001=100)



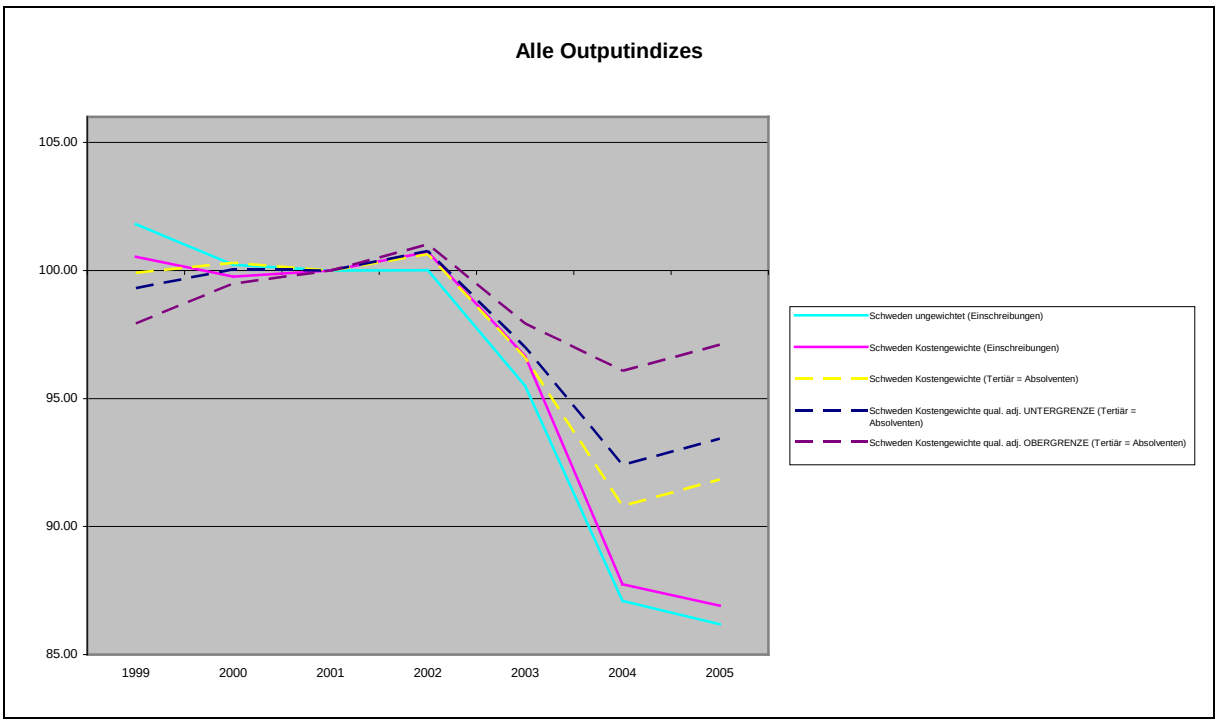
Quelle: UNESCO (2008), eigene Berechnungen.

Abb. A2.2: Output-Entwicklung im Bildungssektor, Italien (Volumen-Index 2001=100)



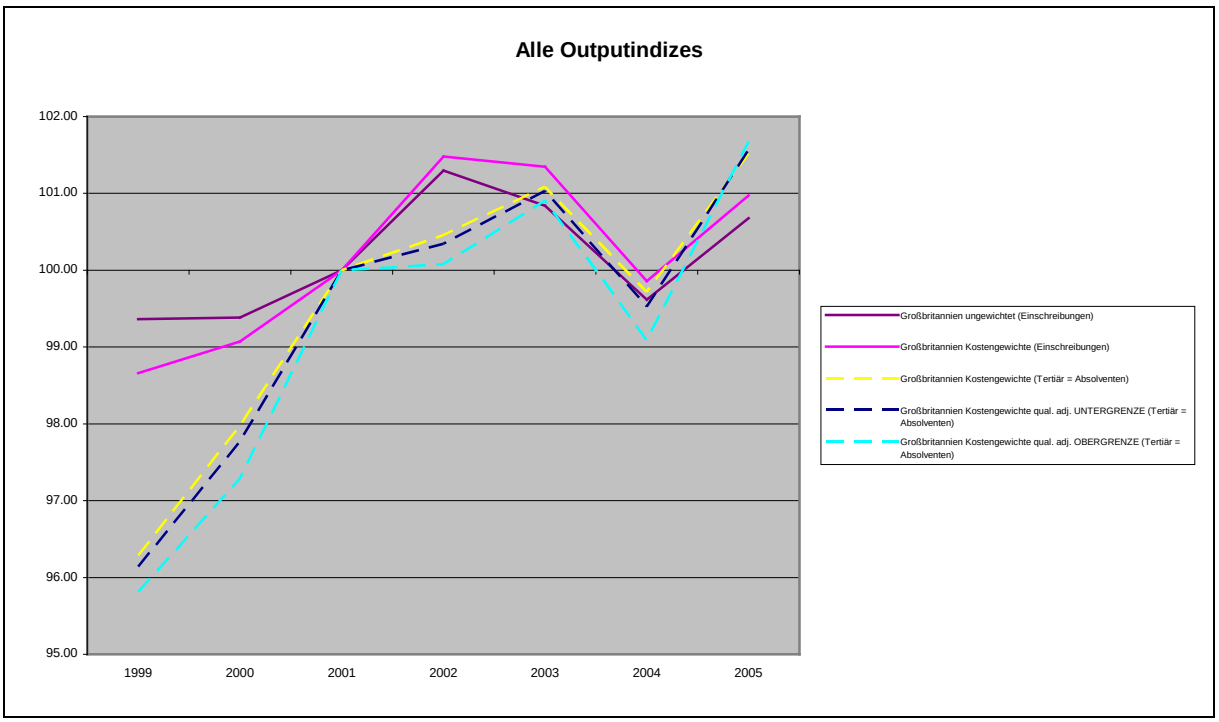
Quelle: UNESCO (2008), eigene Berechnungen.

Abb. A2.3: Output-Entwicklung im Bildungssektor, Schweden (Volumen-Index 2001=100)



Quelle: UNESCO (2008), eigene Berechnungen.

Abb. A2.4: Output-Entwicklung im Bildungssektor, Großbritannien (Volumen-Index 2001=100)



Quelle: UNESCO (2008), eigene Berechnungen.