

Indikatoren der Qualität von Statistiken

Robert Zorzi

Statistiken über wirtschaftliche und gesellschaftliche Strukturen spielen eine bedeutende Rolle bei der Entscheidungsfindung in Politik, Wirtschaft und Gesellschaft. Die Frage nach der Zuverlässigkeit von Statistiken ist in diesem Zusammenhang von zentraler Bedeutung. Eine objektive Einschätzung der statistischen Qualität ist jedoch eine komplexe Aufgabe, insbesondere für die Nutzer von Statistiken, die oft wenig Einblick in die Hintergründe der statistischen Arbeit haben. Mit dem Data Quality Assessment Framework hat der Internationale Währungsfonds einen umfassenden und vor allem standardisierbaren Ansatz zur Beurteilung der statistischen Prozesse und Produkte geschaffen. Konkrete Umsetzungen dieses theoretischen Konzepts im Rahmen der Zahlungsbilanzstatistik durch die Europäische Zentralbank und die Oesterreichische Nationalbank demonstrieren die Nützlichkeit, aber auch die Grenzen von statistischen Qualitätsindikatoren in der Praxis.

1 Rolle von Statistiken und deren Qualität

Statistiken sind nicht nur Werkzeuge der akademischen, makroökonomischen Analyse, sondern stehen zunehmend im Fokus der Märkte und der Tagespolitik. Bei aufmerksamer Verfolgung der Berichte in den Medien kann immer wieder festgestellt werden, dass statistische Indikatoren eine Rolle bei Kursbewegungen auf Wertpapier-, Devisen- und Produktmärkten spielen. Auch politische Diskussions- und Entscheidungsprozesse werden von den als relevant empfundenen statistischen Kennziffern (Maastricht-Kriterien, Wirtschaftswachstum, Arbeitslosenraten etc.) beeinflusst.

In Zeiten der Globalisierung mit zunehmender Verflechtung von Märkten im realwirtschaftlichen und finanzwirtschaftlichen Bereich ist es auch nicht verwunderlich, dass der Bedarf an relevanten, zuverlässigen statistischen Daten steigt. Zunehmende Komplexität der wirtschaftlichen Strukturen geht Hand in Hand mit zunehmendem Wunsch nach Information, Transparenz und Steuerung.

Aufgrund dieser Bedeutung ist es notwendig, die Relevanz und Zuverlässigkeit der im Fokus der Öffentlichkeit stehenden statistischen Daten von einem übergeordneten Aspekt zu betrachten. Dieser Beitrag beleuchtet einige ausgewählte internationale Initiativen, die in den letzten

Jahren gesetzt wurden, um die Beurteilung der Qualität von Statistiken in einem vorgegebenen Rahmen zu ermöglichen. Neben den konzeptionellen Grundlagen, die insbesondere vom Internationalen Währungsfonds (IWF) geschaffen wurden, werden auch erste konkrete Umsetzungen dieser Konzepte im Rahmen der Zahlungsbilanzstatistik durch die Europäische Zentralbank (EZB) und durch die Oesterreichische Nationalbank (OeNB) erläutert.

2 Qualitätskonzept des IWF

2.1 Statistik als Prozess der Datenerstellung

Das *Data Quality Assessment Framework (DQAF)* des IWF aus dem Jahr 2003 ist ein umfassender, formalisierter Ansatz zur Qualitätsbeurteilung von Statistiken. Ähnliche Richtlinien werden von nationalen und internationalen Statistikbehörden schon seit längerem angewandt. Bemerkenswert am Konzept des IWF ist vor allem die ganzheitliche Sichtweise. Nicht nur das statistische Endprodukt und seine Nützlichkeit, sondern der gesamte Prozess der Statistikerstellung wird in die Betrachtung mit einbezogen. Es steht somit das gesamte statistische System, von den rechtlichen und institutionellen Rahmenbedingungen über die Methodologie bis zur Publikationsstrategie, auf dem Prüfstand.

2.2 Voraussetzungen für statistische Qualität

Das DQAF nennt eine Reihe von institutionellen Voraussetzungen beim Statistikproduzenten, die erfüllt sein müssen, bevor man überhaupt von einer tauglichen Statistik sprechen kann. Die wesentlichen Aspekte, die dabei eine Rolle spielen, sind:

- solide gesetzliche Grundlage für die Datenerhebung;
- klare Regelung, welche Institution für die Statistik verantwortlich zeichnet;
- Koordination und reibungsfreier Informationsfluss zwischen den beteiligten Stellen;
- Sicherstellung der Geheimhaltung von nicht anonymisierten Daten;
- ausreichende personelle und technische Ressourcen beim Statistik-ersteller.

2.3 Fünf Dimensionen der statistischen Qualität

Zur eigentlichen Beurteilung der Qualität einer Statistik nennt der IWF-Ansatz fünf übergeordnete Dimensionen, die wiederum durch Elemente näher beschrieben und in weiterer Folge auf Ebene detaillierter Indikatoren quantitativ gemessen bzw. qualitativ beurteilt werden können. Es folgt ein Überblick über diese Dimensionen mit einigen repräsentativen Beispielen für die daran anknüpfenden Elemente und Indikatoren.

2.3.1 Integrität des Statistikerstellers

Objektivität und Unabhängigkeit des Statistikerstellers stehen für den IWF im Zentrum dieser Qualitätsdimension. Unterteilen lässt sich diese Dimension in die Elemente

- Professionalität der Institution;
- Transparenz hinsichtlich statistischer Praktiken und Prozesse;
- Verhaltensregeln für das statistische Personal.

Als Indikator für die Professionalität dient beispielsweise die Unabhängigkeit der jeweiligen Institution bei der Entscheidung über statistische Erhebungsmethoden und -techniken sowie über die Art und Weise der Publikation der Daten. Nur statistische sowie makroökonomische, aber keine politischen Erwägungen sollten hier eine Rolle spielen. Auch sollte jeder privilegierte Zugriff auf die Daten durch Regierungsstellen oder spezielle Interessensgruppen vermieden werden.

2.3.2 Korrekte Anwendung der statistischen Methodologie

Für die meisten Statistiken existieren detaillierte *internationale Standards*¹, die beispielsweise vom IWF, von der OECD oder auf europäischer Ebene von der Europäischen Kommission oder von der EZB festgelegt werden. In manchen Fällen haben diese Standards sogar einen gesetzlichen Status; z. B. hat das Europäische System der Volkswirtschaftlichen Gesamtrechnung in der Europäischen Union (EU) den Rang einer Verordnung.

Die Einhaltung dieser internationalen Richtlinien („standards and good practices“) ist der Kern dieser Dimension. Abweichungen von anerkannten Standards müssen nach Ansicht des IWF zumindest transparent gemacht werden. Die Elemente dieser Dimension sind

- Beachtung der anerkannten volkswirtschaftlichen Konzepte und Klassifikationen;

¹ Definitionen, welches wirtschaftliche Phänomen auf welche Weise in Form von statistischen Aggregaten gemessen werden soll.

- Vollständigkeit der Erhebung (alle relevanten Branchen, Märkte etc.);
 - korrekte Anwendung von Bewertungsrichtlinien.
- Qualität und zeitgerechte Verfügbarkeit der Ausgangsdaten;
 - Abdeckungsgrad der Ausgangsdaten bzw. Repräsentanz der Befragung;
 - Tauglichkeit der statistischen Techniken und Schätzmethoden;
 - regelmäßige Überprüfung von Ausreißern, ungewöhnlichen Trends und scheinbaren oder tatsächlichen Strukturbrüchen;
 - Stabilität der statistischen Zeitreihen im Hinblick auf Datenrevisionen.

Die Beurteilung der Methodologie der Datenerhebung setzt volkswirtschaftliches Grundlagenwissen voraus. Nicht alle theoretisch vorgesehenen statistischen Definitionen sind für das korrekte Endergebnis gleich bedeutsam. Dementsprechend bestimmt die Relevanz eines wirtschaftlichen Phänomens im Gesamtzusammenhang auch die Anwendung von Indikatoren, die sich hinsichtlich der Einhaltung von Definitionen mit diesem Phänomen auseinandersetzen.

Im Hinblick auf die Vollständigkeit einer Erhebung stellt die zugrunde liegende ökonomische Struktur eine natürliche Grenze dar. Eine Volkswirtschaft mit ausgeprägter Präsenz von multinationalen Konzernen und von internationalen Finanzzentren muss anders behandelt werden als eine Volkswirtschaft, die von Klein- und Mittelbetrieben geprägt ist.

Ein häufig angewandter Indikator im Bereich der Bewertungsrichtlinien bei vielen finanzwirtschaftlichen Statistiken ist die Frage, ob die Bewertung zu Marktpreisen erfolgt. Die Anwendung von Nominal-, Buch- oder Bilanzwerten ist üblicherweise nur zulässig, wenn es – etwa mangels eines geeigneten Marktes für ein Produkt oder Finanzinstrument – keine anerkannten, allgemein verfügbaren Marktpreise gibt.

2.3.3 Richtigkeit und Zuverlässigkeit der Datenquellen

Dieser Qualitätsbereich steht meist im Mittelpunkt des Interesses, da hier eine Aussage darüber getroffen werden soll, inwiefern die Statistik *die Realität tatsächlich widerspiegelt*. Dieser Bereich umfasst die Elemente

Aussagen über Ursprung, Definition, Umfang und Gliederungstiefe lassen qualitative Schlüsse über die Güte der verwendeten Ausgangsdaten zu. Ist die Quelle der Daten beispielsweise die Bilanz eines Unternehmens, so lässt sich meist anhand der jeweils üblichen Buchhaltungsvorschriften ableiten, in welchem Ausmaß die Daten den statistisch benötigten Definitionen tatsächlich entsprechen. Statistiker sehen sich dabei sehr oft mit der Herausforderung konfrontiert, die Brücke von den vorhandenen betriebswirtschaftlichen Daten zu den geforderten volkswirtschaftlichen Abgrenzungen zu schlagen.

Die Erhebungsquelle beeinflusst in hohem Maß auch die zeitgerechte Verfügbarkeit der Ausgangsdaten. Was zeitgerechte Verfügbarkeit bedeutet, lässt sich jeweils aus dem geforderten Publikationstermin des fertigen Statistikprodukts ableiten. Besonders für die erste Version einer Statistik ist es entscheidend, welcher Anteil der erforderlichen Ausgangsdaten zum Publikationszeitpunkt bereits vorhanden war.

Auch das Erhebungsdesign bildet eine wesentliche Grundlage für die quantitative Beurteilung der statistischen Qualität. Die maßgeblichen Faktoren dafür sind die Art und der

Umfang der Stichprobe, falls eine Vollerhebung, wie in den meisten Fällen, nicht möglich oder sinnvoll ist. Bei Konzentrationsstichproben stehen die Höhe der Schwellenwerte bei der Auswahl von Meldern und die daraus ableitbaren Abdeckungsgrade im Zentrum des Interesses. Beispielsweise empfiehlt der IWF für die meisten Statistiken einen Abdeckungsgrad von 95% bis 99%. Bei Zufallsstichproben hängt die Qualität nicht nur von der Anzahl der Befragten, sondern auch davon ab, ob eine einfache oder geschichtete Stichprobe verwendet wurde. Konfidenzintervalle und Stichprobenfehler sind wesentliche Messgrößen in diesem Zusammenhang.

Viele statistische Aggregate werden in regelmäßigen Abständen revidiert und neu veröffentlicht, da zusätzliche Daten verfügbar wurden bzw. da Fehler und Ungenauigkeiten in den ursprünglichen Daten oder in den Berechnungsmethoden nachträglich behoben wurden. Aus dem durchschnittlichen Ausmaß von Revisionen lässt sich mit verschiedenen Indikatoren die Stabilität und somit die Zuverlässigkeit der Erstpublikation ableiten.

2.3.4 Zweckmäßigkeit der Statistik

Diese Dimension misst gewissermaßen die *Nützlichkeit bzw. Verwendbarkeit der Statistiken*. Eine Statistik ist nur dann sinnvoll, wenn die Daten auch für Analysen, Studien und Entscheidungsfindungen eingesetzt werden können. Messen kann man dies anhand der Elemente

- benutzeradäquate Periodizität und zeitgerechte Verfügbarkeit der Statistik;
- interne und externe Konsistenz der Daten.

Hinsichtlich der zeitlichen Komponente stellt der IWF die Frage, ob die Statistik entsprechend den Nutzeranforderungen genügend schnell und in der richtigen Frequenz vorhanden ist. Hier führt der IWF das Beispiel an, dass Daten, die immer erst nach einem wesentlichen Entscheidungsprozess verfügbar werden, wenig zweckmäßig sind. Verlangt die Meinungsbildung eine Zeitreihe auf vierteljährlicher Basis, dann ist die Erstellung von nur jährlichen Summen ebenfalls suboptimal.

Eine anerkannte Referenz hinsichtlich Periodizität und rechtzeitiger Verfügbarkeit von wichtigen volkswirtschaftlichen Kenngrößen ist der *Special Data Dissemination Standard (SDDS)* des IWF.

Die Konsistenz einer Statistik ist ein Qualitätselement, das sich sehr gut für eine quantitative Beurteilung eignet. Von interner Konsistenz spricht man in diesem Zusammenhang, wenn die Statistik in sich schlüssig ist. Eine auf alle Statistiken anwendbare Form der internen Konsistenz besteht in der Vergleichbarkeit der Daten im Zeitverlauf. Brüche in den Zeitreihen sollten nicht zu häufig auftreten und für den Benutzer gut dokumentiert und nachvollziehbar sein. In Finanzstatistiken ist die Konsistenz zwischen Anfangsstand, Transaktionen, Bewertungsänderungen und Endstand einer Periode eine häufig verwendete Konsistenzprüfung.

Externe Konsistenz lässt sich immer dann messen, wenn es mehrere verschiedene statistische Quellen gibt, die gänzlich oder zumindest in Teilbereichen dasselbe Phänomen beschreiben. Die Übereinstimmung zwischen diesen Quellen kann als absolute oder prozentuelle durchschnittliche Differenz dargestellt werden.

2.3.5 Verfügbarkeit und Zugänglichkeit der Statistik

Zu guter Letzt muss die statistische Publikation für den Nutzer leicht zugänglich und verständlich sein. Dies umfasst die Elemente

- regelmäßige Publikation in geeigneten Medien und Formaten;
- Verfügbarkeit von Metadaten zu Ursprung und Inhalt der Statistiken;
- Kommentierung der Ergebnisse (Presseaussendungen, Studien);
- Unterstützung der Nutzer bei Anfragen (z. B. Statistik Hotline).

Indikatoren für diese Qualitätselemente sind beispielsweise die Veröffentlichung in elektronischer Form im Internet, die einfache Lesbarkeit und Auswertbarkeit von Tabellen und Grafiken, die Gleichbehandlung aller potenziellen Nutzer (gleichzeitiger Datenzugang zu den gleichen Bedingungen), die regelmäßige Veröffentlichung von Presseaussendungen und weiterführenden Studien, die Veröffentlichung und Einhaltung eines Publikationskalenders und die Schaffung einer zentralen Ansprechstelle für Anfragen.

2.3.6 Messung statistischer Qualität im Rahmen der fünf Qualitätsdimensionen

Wie die angeführten Beispiele zeigen, lässt sich jede Dimension durch eine Vielzahl von Indikatoren beschreiben, die sich einer prozessorientierten Beurteilung bzw. Messung erschließen.

Die Empfehlung des IWF zur praktischen Umsetzung beschränkt sich im ersten Schritt auf einen qualitativen Ansatz. Das heißt, nach der Auswahl der geeigneten Indikatoren zu allen Elementen und Dimensionen erfolgt die Beurteilung eines jeden Indikators auf einer vierteiligen Skala. Diese Skala erstreckt sich von „Bedin-

gung erfüllt“ über „Bedingung größtenteils erfüllt“ und „Bedingung teilweise erfüllt“ bis zu „Bedingung nicht erfüllt“.

2.4 Anmerkungen zur Relevanz des Qualitätskonzepts des IWF

Mit dem DQAF erhebt der IWF den Anspruch, ein Rahmenwerk anzubieten, das Statistiken und statistische Praktiken international vergleichbar macht. Durch die *Standardisierbarkeit und Wiederholbarkeit von Vergleichen* wird man diesem Anspruch durchaus gerecht. Der IWF vermeidet überdies jede Reihung der Qualitätsdimensionen nach Priorität. Das heißt, kein Element und kein Indikator ist von vornherein wichtiger als die anderen Qualitätsaspekte.

Ein weiterer begrüßenswerter Aspekt des DQAF besteht darin, dass zumindest theoretisch die *Abhängigkeiten* zwischen den Qualitätsdimensionen berücksichtigt werden, um eine faire Qualitätsbeurteilung sicherzustellen.

Typisches Beispiel für eine solche Abhängigkeit ist der Zusammenhang zwischen schneller Verfügbarkeit und Richtigkeit bzw. Zuverlässigkeit der Daten. Je schneller eine Statistik verfügbar gemacht wird, desto schwieriger ist es, die Zuverlässigkeit der Daten hinsichtlich Erschließung und Prüfung aller Ausgangsdaten sicherzustellen.

Der sehr ambitionierte Ansatz, der dem DQAF zugrunde liegt, hat jedoch seinen Preis. In der praktischen Umsetzung stößt man – wie später näher ausgeführt wird – sehr schnell an die operativen Grenzen. Die qualitative Natur der meisten Indikatoren impliziert eine mitunter schwer nachvollziehbare, subjektive Beurteilung der Qualität.

Die Qualitätsdimension zur Methodologie ist in dieser Hinsicht besonders problematisch zu sehen. Es wird zwar angegeben, ob die internationalen Standards eingehalten werden, es bleibt für den Nutzer der Daten aber offen, welche praktische Relevanz etwaige Abweichungen tatsächlich haben bzw. es wird nicht hinterfragt, ob der jeweilige Standard aus Nutzersicht wirklich nachvollziehbar und sinnvoll ist.

3 Qualitätsbericht der EZB für die Zahlungsbilanz des Euroraums

3.1 Task Force on Quality

Im Rahmen einer Task Force des Europäischen Systems der Zentralbanken (ESZB) unter Mitwirkung von Eurostat wurde im Jahr 2003 der Versuch gestartet, die Richtlinien des IWF so weit als möglich am Beispiel eines konkreten statistischen Produkts in die Praxis umzusetzen. Ein wesentliches Ziel der Arbeiten war die Nachvollziehbarkeit der Qualitätsmessungen. Dies führte dazu, dass der Schwerpunkt auf quantitativ messbare Qualitätskriterien gelegt wurde.

Dazu zählen vor allem die Stabilität von statistischen Zeitreihen im Hinblick auf die „Anfälligkeit“ für Revisionen und die Konsistenz der Daten innerhalb der Statistik und zu anderen statistischen Quellen. Für diese Indikatoren wurden zahlreiche mathematische Formeln entwickelt. Besonders im Bereich der Revisionsindikatoren legte die EZB Wert auf intensive Analysen von Ursachen und Mustern von Datenrevisionen.

Die Qualitätsindikatoren, die von der Task Force on Quality entwickelt wurden, gingen somit eher in die Richtung eines am statistischen Endprodukt orientierten, quantitativen Ansatzes im Gegensatz zu dem ganz-

heitlichen, prozessorientierten Konzept des IWF. Dies implizierte auch eine Reihung der Qualitätsdimensionen nach Prioritäten, was ebenfalls eine gewisse Abweichung von den Intentionen des DQAF bedeutete.

3.2 Erste Publikation der EZB zur Qualität der Euroraum-Zahlungsbilanz

Im Jänner 2005 veröffentlichte die EZB einen Bericht zur Qualität der Daten der Zahlungsbilanz und der Internationalen Vermögensposition des Euroraums. Dieser Bericht ist als Pilotversuch für die Statistiken des Euroraums zu sehen und soll in ähnlicher Form jährlich verfügbar gemacht werden. Gegenstand der Qualitätsanalyse waren die Daten des Euroraums von 1999 bis 2003.

Die Untersuchungen der EZB konzentrierten sich im konkreten Fall auf die im Folgenden beschriebenen Qualitätselemente, die eine Untermenge des DQAF des IWF darstellen.

3.2.1 Qualität der Methodologie der Zahlungsbilanz

Besonders günstig fiel das Urteil der EZB hinsichtlich der angewandten Methodologie bei der Erstellung der Zahlungsbilanz aus. Für die Länder der EU existiert seit fast zehn Jahren ein umfassendes Methodologiekompilium (*EU b.o.p. and i.i.p. statistical methods*), das jedes Jahr im November von der EZB in einer aktualisierten Version publiziert wird. In dieser Publikation werden die Erhebungssysteme und -methoden aller Länder detailliert beschrieben, wobei auch auf Abweichungen der statistischen Praxis von internationalen Standards hingewiesen wird. Im Zeitverlauf konnte anhand dieser Dokumentation nachgewiesen werden, dass seit Beginn der Währungsunion in den ein-

zelen Ländern laufend Maßnahmen gesetzt wurden, um bestehende methodologische Lücken zu schließen.

3.2.2 Zeitgerechte Verfügbarkeit der Zahlungsbilanz

Auch hinsichtlich der zeitgerechten Verfügbarkeit der Ergebnisse konnte belegt werden, dass alle Publikationstermine eingehalten wurden. Im Bericht der EZB wird besonders darauf hingewiesen, dass für den Euroraum monatliche Zahlungsbilanzdaten nach bereits ca. sieben Wochen veröffentlicht werden. Dies ist ein im internationalen Vergleich sehr guter Wert.

3.2.3 Stabilität der Zahlungsbilanz

Zur Beurteilung der Stabilität der Zeitreihen (Umfang der Revisionen) wurden drei quantitative Indikatoren bzw. Formeln verwendet.

Die Kennzahl *Directional Reliability* gibt Auskunft darüber, ob eine Statistik schon in den ersten, meist noch vorläufigen Daten die Trends (Anstieg, Rückgang etc.) richtig widerspiegelt. Im Fall der Zahlungsbilanz wurden diesbezüglich nur im Bereich der Direktinvestitionen Schwächen geortet, die auf enorme Unsicherheiten bei der Schätzung reinvestierter Gewinne (d. h. erwirtschafteter, aber nicht ausgeschütteter Dividenden) auf monatlicher Ebene zurückzuführen sind. Im internationalen Vergleich dürfte der Euroraum hier aber dennoch gut liegen, da in manchen Ländern derartige Schätzungen noch nicht einmal Teil der Zahlungsbilanzmethodologie sind.

Der Indikator *Mean Absolute Percentage Error (MAPE)* wurde in erster Linie für die Beurteilung der Bruttoströme der Leistungsbilanz entwickelt und ist ein Maß für die durchschnittliche Differenz zwischen Erstpublika-

tion und endgültigen Daten über mehrere Berichtsperioden. Lediglich zu Beginn der Währungsunion im Jahr 1999 ergaben sich durch die damals notwendigen Umstellungen bzw. Initialisierungen der statistischen Berechnungsmethoden für den Euroraum ungewöhnlich hohe Revisionen.

Der *Root Mean Square Relative Error (RMSRE)* wird auf Nettotransaktionen in der Kapitalbilanz angewandt und misst die Differenz zwischen vorläufigen und endgültigen Daten im Verhältnis zur allgemeinen Volatilität einer Zeitreihe. Diese etwas komplexere Methode wurde entwickelt, da bei sehr stark schwankenden Zeitreihen mit von Periode zu Periode unterschiedlichen Vorzeichen einfache Durchschnittsindikatoren wie MAPE wenig aussagekräftig sind. Analog zu den Ergebnissen der „Directional Reliability“ zeigten sich bei der Anwendung dieses Indikators größere Diskrepanzen im Bereich der Direktinvestitionen aufgrund der schwierigen Datenlage bei der Schätzung monatlicher reinvestierter Gewinne.

3.2.4 Konsistenz der Zahlungsbilanzdaten

Im Bereich der Konsistenz der Zahlungsbilanzdaten wurden drei Indikatoren zur Qualitätsbeurteilung herangezogen.

Als Maßstab für die interne Konsistenz bot sich bei der Zahlungsbilanz die Auswertung der *Statistischen Differenz* an. Die Berechnung ergab eine durchschnittliche Höhe der Statistischen Differenz von 15 Mrd EUR pro Monat. Dies entspricht ca. 5% der durchschnittlichen Bruttoströme der Leistungsbilanz. Im Zeitverlauf von 1999 bis 2003 zeigt sich eine leichte Abnahme der Statistischen Differenz.

Die externe Konsistenz der Zahlungsbilanz des Euroraums wurde anhand von Vergleichen mit den *nach EU-Standards erhobenen Daten zum Außenhandel* sowie mit den aus Beständen der *EZB-Monetärstatistik* abgeleiteten grenzüberschreitenden Kapitaltransaktionen der Monetären Finanzinstitute geprüft.

Im Fall der Daten zum Warenverkehr wurde, da es sich um Bruttoströme handelt, der Durchschnitt der absoluten Differenzen (analog MAPE) ermittelt. Es zeigte sich eine im Zeitverlauf konstant bleibende Differenz zwischen Waren laut Zahlungsbilanz und laut Außenhandel von ca. 2% auf der Exportseite und von ca. 5% auf der Importseite. Dieses Ergebnis entspricht laut EZB-Bericht durchaus den Erwartungen an die Qualität der Zahlungsbilanzdaten. Güterströme laut Außenhandel und laut Zahlungsbilanz unterscheiden sich auf methodologischer Ebene in einigen wesentlichen Punkten (z. B. Bewertungen der Warentransaktionen inklusive oder exklusive Fracht- und Versicherungskosten), woraus eine relativ konstante absolute Differenz bei gleichzeitig hoher Übereinstimmung in den Wachstumsraten resultieren sollte.

Der Vergleich mit den Daten der *EZB-Monetärstatistik* erfolgte anhand einer Variante des RMSRE, da es sich um stark volatile Nettokapitalströme handelt. Eine durchschnittliche Abweichung zwischen 10% und 15% wurde ermittelt. Dieser Wert ist etwas höher als die bekannten methodologischen Unterschiede zwischen den beiden Statistiken erwarten ließen. Andererseits ist absehbar, dass sich dieses Verhältnis in naher Zukunft verbessern wird, da viele Länder dazu übergehen, die EZB-Monetärstatistik als direkte und einzige Quelle auch

für die Zahlungsbilanz und für die Internationale Vermögensposition zu verwenden (in Österreich ab Jänner 2006).

3.2.5 Zusammenfassendes Qualitätsurteil zur Zahlungsbilanz

Insgesamt kommt die EZB zu dem Schluss, dass die Euroraum-Zahlungsbilanz auf Basis der verwendeten Qualitätskriterien als glaubwürdig und zuverlässig eingestuft werden kann. Insbesondere zeigt sich seit den ersten Publikationen im Jahr 1999 eine kontinuierliche Verbesserung der Situation, die auf die bessere Routine sowie auf zahlreiche individuelle Maßnahmen der Mitgliedstaaten im Zusammenhang mit der Verbesserung der Praxis der Datenerhebung zurückzuführen ist.

3.3 Anmerkungen zur Relevanz des Qualitätsberichts der EZB

Aus methodischer Sicht ist der Qualitätsbericht der EZB zweifellos ein bedeutender Schritt in die richtige Richtung, spiegelt aber auch die Probleme bei der praktischen Umsetzung der IWF-Kriterien sowie die dabei entstehenden Risiken wider. Analog den Ergebnissen der zuvor erwähnten Task Force on Quality verzichtete die EZB darauf, alle vom IWF vorgeschlagenen Kriterien zu durchleuchten und legte den Schwerpunkt auf jene Dimensionen, die aus Sicht der EZB am ehesten für eine dem externen Benutzer verständliche Implementierung geeignet erschienen.

Dementsprechend zeigt sich eine eher *einseitige Ausrichtung der EZB auf quantitative Kriterien und eine Vernachlässigung der anderen Dimensionen*, die schwieriger zu beurteilen, aber nicht weniger relevant für das Gesamtbild sind. Fast 80% des EZB-Berichts beschäftigen sich mit den

beiden Kriterien Stabilität und Konsistenz der Zeitreihen. Die quantitativen Analysen dazu sind sehr aufschlussreich und auch die daraus abgeleiteten Schlussfolgerungen sind durchaus fair gegenüber den nationalen Statistikproduzenten. Allerdings lassen sich viele dieser quantitativen Indikatoren aus Kundensicht nur sehr schwer interpretieren. Beispielsweise bleibt offen, ob die ermittelten Durchschnittswerte für die Statistische Differenz nun als zufrieden stellend oder als problematisch anzusehen sind. Generell fehlen die Vergleichsmöglichkeiten (*Benchmarks*) für eine sinnvolle Einordnung dieser Qualitätsindikatoren. Eine bessere Unterstützung bei der Interpretation der Indikatoren für den mit der Materie nicht vertrauten Leser hätte die Akzeptanz für diesen Qualitätsbericht sicher erhöht.

Weiters besteht die Gefahr, dass die Leser eines solchen Berichts zum Schluss kommen, letztendlich seien Stabilität und Konsistenz die einzig wichtigen Kriterien zur Qualitätsbeurteilung. Eine Konsequenz aus dieser einseitigen Sichtweise könnte langfristig darin bestehen, dass sich die Statistikproduzenten ebenfalls nur noch auf die Erfüllung der scheinbar bevorzugten, quantitativen Qualitätskriterien konzentrieren, um für die externen Nutzer ein gutes Bild abzugeben.

Vor dem Hintergrund enger Publikationstermine sind beispielsweise Revisionen aufgrund von Korrekturen und verspätet eingelangten Meldungen unvermeidlich. Das Fehlen solcher Revisionen ist somit nicht unbedingt ein sicheres Zeichen dafür, dass die einmal veröffentlichte Statistik tatsächlich richtig ist. Im Endeffekt sind Revisionen immer eine Verbesserung der Daten, deren Unterdrückung nicht das Ziel der statistischen

Arbeit sein darf. Es liegt immer in der Verantwortung des Statistikerstellers, eine Ausgewogenheit zwischen Häufigkeit und Ausmaß von Revisionen zu finden.

Auch die externe Konsistenz von Statistiken ist als Qualitätsindikator ein zweischneidiges Schwert. Treten Inkonsistenzen zwischen zwei statistischen Quellen auf, dann ist der Statistikkunde sicher nicht in der Lage, auf Basis von reinen Konsistenzindikatoren zu entscheiden, welche Quelle für seine Zwecke nun die zuverlässigere oder relevantere ist. Nur die Produzenten der Statistiken können die nötigen Evaluierungen durchführen und dem Kunden die entsprechenden Hinweise geben. Werden keine solchen Maßnahmen getroffen, dann ist damit zu rechnen, dass eine der Quellen früher oder später angepasst oder gänzlich aufgelassen wird. Damit wird zwar Konsistenz geschaffen, aber der Konsistenzindikator zur Qualitätsbeurteilung wird mangels einer echten Vergleichsmöglichkeit obsolet.

4 Qualitätskonzept des IWF und das neue System für Außenwirtschaftsstatistiken in Österreich

Bis Ende 2005 beruht die Erstellung der österreichischen Zahlungsbilanz in erster Linie auf Meldungen der Banken über ihre eigenen *Auslandszahlungen* und jene ihrer Kunden mit dem Ausland. Aufgrund von Änderungen in den Zahlungsverkehrssystemen in Folge der Einführung des Euro und aufgrund der zunehmenden Komplexität der Finanzierungs- und Abrechnungsströme multinationaler Unternehmen ist es mittelfristig nicht mehr möglich, auf dieser Datenbasis zuverlässige Zahlungsbilanzstatistiken zu erstellen. Nicht nur Österreich, sondern auch zahlreiche andere Staa-

ten in der EU stellen die Außenwirtschaftsstatistik deshalb auf *Direkterhebungssysteme* um, die keinen unmittelbaren Bezug zum Zahlungsverkehr mehr haben. In Österreich wird die Umstellung per 1. Jänner 2006 stattfinden.

Dieser Umstellungsprozess war für die OeNB als Institution, die gesetzlich für die Erstellung der österreichischen Zahlungsbilanz verantwortlich ist, die beste Gelegenheit, die beschriebenen Qualitätskriterien in ihrer Gesamtheit beim Entwurf eines neuen Systems zu berücksichtigen.

4.1 Entscheidung zur Umstellung der Zahlungsbilanz in Österreich

Als im Jahr 2002 die Umstellung des Erhebungssystems für die Zahlungsbilanz initiiert wurde, basierte dies auf konkreten Qualitätsüberlegungen.

Die *Schwellenwerte* für statistische Meldungen im Rahmen des Auslandszahlungsverkehrs können seit 2003 nicht mehr autonom von der OeNB festgelegt werden, sondern sind durch EU-Recht geregelt. Die Höhe dieser Meldeschwellen liegt bereits seit Ende 2003 am oberen Limit, um noch einen ausreichenden Abdeckungsgrad für die Statistik zu erreichen. Eine weitere Anhebung wurde für 2006 angekündigt. In den von der OeNB durchgeführten Hochrechnungen wurde deutlich, dass spätestens mit diesem Schritt keine ausreichende *Richtigkeit* und *Zuverlässigkeit* mehr erreicht werden könnte.

Überdies hätte die steigende Anzahl an gemeldeten Einzeltransaktionen und die zunehmende Komplexität der zugrunde liegenden Wirtschaftsabläufe zur Folge, dass es mittelfristig nicht mehr möglich sein würde, in der geforderten Zeit mit vertretbaren Kosten alle nötigen Qualitätsprüfun-

gen und statistisch erforderlichen Korrekturen bis zur Produktion des Endergebnisses durchzuführen. Somit wären sowohl die *zeitgerechte Verfügbarkeit* als auch die *Stabilität* der Daten mittelfristig ernsthaft gefährdet gewesen.

Verschärft wurde diese Situation dadurch, dass sowohl durch Rechtsakte der EZB als auch durch Verordnungen der EU die Anforderungen an Richtigkeit, Zuverlässigkeit und zeitgerechte Verfügbarkeit von Zahlungsbilanzdaten in den letzten Jahren enorm gestiegen sind.

Die OeNB befand sich somit als Statistikproduzent in einer Schere zwischen steigenden Anforderungen der Nutzer der Daten und einem Verlust der Qualität der bisher verwendeten primären Datenbasis. Die Entscheidung für eine grundlegende Umstellung der Erhebungsmethode nach dem Vorbild anderer EU-Mitgliedstaaten war somit unvermeidlich.

4.2 Aufbau des Systems für Außenwirtschaftsstatistiken in Österreich auf Basis der Qualitätsdimensionen

4.2.1 Voraussetzungen für statistische Qualität

Wie im DQAF des IWF vorgesehen, wurden als Erstes die notwendigen rechtlichen und organisatorischen Weichenstellungen für ein neues Zahlungsbilanzsystem in Angriff genommen.

Das neue *Devisengesetz 2004* und die darauf aufbauenden neuen Meldeverordnungen der OeNB bilden auch in Zukunft eine vorbildliche rechtliche Grundlage für die Erstellung der Zahlungsbilanzstatistik für Österreich. Insbesondere ermöglichte die neue Rechtslage der OeNB den Zugriff auf andere statistische und administrative Daten und somit einen

verbesserten Informationsfluss. Auch die für die statistische Praxis wichtige Geheimhaltung der erhobenen Ausgangsdaten wird in diesem Gesetz klar geregelt.

Auf organisatorischer bzw. institutioneller Ebene wurde eine intensive *Zusammenarbeit mit Statistik Austria* in Form eines Kooperationsrahmenvertrags ins Leben gerufen. Eines der wichtigsten Vorhaben im Rahmen dieser Kooperation ist die Umsetzung des neuen Erhebungssystems zu Außenwirtschaftsstatistiken. Um die Expertise der jeweiligen Institution zu nutzen, wurde vereinbart, dass sich die OeNB auf den finanzwirtschaftlichen Bereich konzentriert, während Statistik Austria große Teile der realwirtschaftlichen Erhebungen (vor allem Dienstleistungen) im Auftrag der OeNB durchführen wird. Im Rahmen von konkreten Verträgen erfolgte eine detaillierte Abgrenzung der Aufgabenbereiche mit klaren Regeln für Zuständigkeiten, Informationsflüsse und Ressourceneinsatz.

4.2.2 Integrität und Transparenz der Statistikersteller

Im Zusammenhang mit dem Umstellungsprozess wurde versucht, durch frühzeitige und breite Information der betroffenen Kooperationspartner (Banken, Unternehmen, Interessensvertretungen, staatliche Stellen etc.) Irritationen zu vermeiden und rechtzeitig die nötige Akzeptanz für dieses Projekt zu erreichen. Die OeNB kommunizierte sowohl die Gründe für die anstehende Umstellung als auch die Inhalte des neuen Systems. Zu diesem Zweck wurden das Internet (www.zahlungsbilanz.oenb.at), die Printmedien, Broschüren und fachspezifische Publikationen als Kommunikations-schiene intensiv genutzt. Weiters wurden zahlreiche Informationsver-

anstaltungen (*Road Shows*) für alle Betroffenen sowie spezielle *Arbeitsgruppen* mit einzelnen Melderkreisen organisiert.

4.2.3 Korrekte Methodologie

Jeder Veränderungsprozess ist eine Chance zur Verbesserung der Gesamtsituation. Dies galt im konkreten Fall auch für die korrekte Anwendung der statistischen Methodologie. Auch in der österreichischen Zahlungsbilanz gab es methodologische Verbesserungspotenziale, die im bisher angewandten System aus technischen oder wirtschaftlichen Gründen ohne enormen Zusatzaufwand nicht ausgeschöpft werden konnten. Dies betraf vor allem die Trennschärfe von Detailaggregaten.

Es war deshalb ein wesentliches Ziel des neuen Systems, solche Lücken so weit als möglich zu schließen. Besonders erfolgreich waren diese Bemühungen hinsichtlich der statistischen Erhebungen zu grenzüberschreitenden Versicherungsleistungen sowie zu Handelskrediten im Rahmen von Konzernbeziehungen.

4.2.4 Richtigkeit und Zuverlässigkeit der Daten

Im Hinblick auf die Richtigkeit der künftig erhobenen Daten wurden die zukünftigen Meldeschwellenwerte mit besonderer Sorgfalt festgelegt. Ziel war es, ein ausgewogenes Verhältnis zwischen Abdeckungsgraden und Anzahl der Melder zu erreichen. Gemäß den internationalen Standards wurde jeweils ein Abdeckungsgrad von 95% angestrebt. Ein wesentliches Ziel war es aber auch, die Belastung der Melder möglichst gering zu halten. Als Basis für diese Auswertungen dienten sowohl die Daten aus dem derzeit noch bestehenden Erhebungssystem als auch externe Datenquellen

(z. B. Register). Im Fall der Dienstleistungen wurde in Zusammenarbeit mit Statistik Austria eine eigene Vorerhebung bei 32.000 Unternehmen (quasi als *population survey*) durchgeführt.

4.2.5 Zweckmäßigkeit der Daten

Da die zeitgerechte Publikation der Daten ohnehin auf verbindlichen Vorgaben (z. B. SDDS des IWF) beruht, lag der Schwerpunkt der Arbeiten im Rahmen des neuen Systems auf der *Konsistenz* der Daten. Die Konsistenz zum Außenhandel sowie die Konsistenz zu den Daten der EZB-Monetärstatistik, die auch die EZB in ihrem Qualitätsbericht besonders genau untersuchte, standen auch beim Projekt der OeNB im Mittelpunkt. Es wurde in beiden Fällen das einfachste Mittel gewählt, um hier in Zukunft volle Konsistenz herzustellen: Der Verzicht auf eigene Erhebungen und die unmittelbare Verwendung der Primärstatistiken in der eigenen Statistik. Dies verringert auch den Aufwand, sowohl für die Melder als auch für die OeNB. Qualitätsindikatoren zur externen Konsistenz lassen sich jedoch in Zukunft nicht mehr ableiten.

5 Schlussbemerkungen

Theoretische Konzepte zur Qualität von Statistiken, wie sie der IWF entwickelt hat, sind noch weit davon entfernt, ausgereift und international voll akzeptiert zu sein. Die erwähnten Beispiele der konkreten Anwendung auf europäischer und auf österreichischer Ebene zeigen jedoch, dass diese Ideen durchaus in der konkreten statistischen Arbeit von Nutzen sein können.

Einige besonders wichtige Fragen sind in diesem Zusammenhang aus Sicht des Verfassers auf internationaler Ebene noch nicht genügend berücksichtigt worden.

5.1 Abhängigkeiten zwischen Elementen der Qualität

Es wurde bereits erwähnt, dass zwischen einigen Qualitätsdimensionen komplexe Abhängigkeiten bestehen. Beispielsweise führt eine sehr schnelle Verfügbarkeit der Daten (*timeliness*) zwangsläufig zu Problemen bei der Richtigkeit und Zuverlässigkeit (z. B. *Stabilität*) der Daten. Das heißt, der Versuch, ein Qualitätskriterium möglichst gut zu erfüllen, hat oft zur Konsequenz, dass bei anderen Kriterien nicht mehr dieselbe, hohe Qualitätsstufe erreicht werden kann.

Es ist bisher noch zu selten die Frage gestellt worden, welches Qualitätskriterium für die Abnehmer der Daten wichtiger ist, wobei es oft *verschiedene Abnehmergruppen mit unterschiedlichen Prioritäten* gibt. Die Notwendigkeit genauerer Untersuchungen über die Wünsche dieser Nutzer der Statistiken wurde noch nicht ausreichend berücksichtigt.

5.2 Kosten einer Statistik

Wer einen Bedarf an statistischen Daten hat, hat das legitime Interesse, diese Daten in der gewünschten Detaillierung, Genauigkeit und Rechtzeitigkeit kostenlos zu beziehen. Man sollte jedoch nie vergessen, dass die Erstellung von Statistiken mit einem nicht zu unterschätzenden Aufwand verbunden ist.

In dieser Hinsicht ist einerseits der Aufwand jener Institution zu berücksichtigen, die diese Daten auf gesetzlicher Grundlage zu erstellen hat. Diese Institutionen befinden sich meist in einem mehr oder minder starken Nahverhältnis zur öffentlichen Hand. Es ist somit letztendlich immer ein öffentlicher Aufwand, der hier in Kauf genommen werden muss. Umgekehrt stehen die Nutzer der Daten zu einem großen Teil ebenfalls im Ein-

flussbereich der öffentlichen Verwaltung. Konflikte zwischen den Ansprüchen der Nutzer der Daten und den finanziellen und personellen Möglichkeiten der Produzenten sind trotzdem (oder gerade deswegen) nicht selten.

Auf der anderen Seite steht der noch größere Aufwand, den die „Zulieferer“ des Statistikproduzenten auf sich nehmen müssen, nämlich die so genannten Melder. Es handelt sich dabei meist um finanzielle und nicht-finanzielle Unternehmen, aber auch um andere wirtschaftliche Einheiten und manchmal sogar um Privatpersonen. Diese Melder sind oft nur indirekte Nutzer der statistischen Daten. Es lässt sich zwar überzeugend argumentieren, dass es für Österreich als hoch entwickelte Volkswirtschaft mit sehr starker Einbindung in das globale Wirtschafts- und Finanzsystem gesamtwirtschaftlich unverzichtbar ist, den Märkten und Entscheidungsträgern ausreichende statistische Kennzahlen zur Verfügung zu stellen; diese Sichtweise einem Melder, der einen entsprechenden Kostenaufwand zur Ermittlung der von ihm abverlangten Daten tragen muss, zu vermitteln, erfordert aber viel Fingerspitzengefühl.

5.3 Kommunikation als Element der Qualität

Für die OeNB war es schon bisher selbstverständlich, die von ihr produzierten Statistiken einer breiten Öffentlichkeit möglichst zweckmäßig zur Verfügung zu stellen und entsprechende Unterstützung für weitergehende Anfragen anzubieten. In dieser Hinsicht sind die vor kurzem neu gestaltete Homepage (www.oenb.at) mit einem eigenen Bereich für Statistiken bzw. statistische Publikationen sowie die seit längerem etablierte Sta-

tistik Hotline der OeNB (01/40420-5555) hervorzuheben.

Neben diesen Informationsangeboten, die zum Standard eines Statistikerstellers gehören sollten, ist es aber auch erforderlich, die Daten selbst den potenziellen Nutzern näher zu bringen. Dabei ist zu bedenken, dass Daten nicht für sich alleine sprechen. Abgesehen von der regelmäßigen Kommentierung und Analyse der Statistiken in Form von Presseaussendungen, Berichten und Studien, ist es auch erforderlich, mehr als nur reine Qualitätsindikatoren in Form von Kennzahlen (z. B. MAPE, Konfidenzintervalle, Abdeckungsgrade etc.) zur Verfügung zu stellen. Die richtige Interpretation der Indikatoren zur Qualität von Statistiken erfordert ein hohes Maß an Begleitung und Erklärung durch den Statistikersteller. Der Statistikproduzent kennt die Stärken und Schwächen seines statistischen Systems selbst am besten und muss bemüht sein, die relevantesten Aspekte für die Nutzer der Daten in der richtigen Gewichtung transparent zu machen.

5.4 Resümee: Akzeptanz bestimmt die Qualität des Outputs

Es ist legitim und notwendig, Anforderungen an die Qualität von Statistiken zu stellen und diese auch systematisch zu überprüfen. Das Bild ist jedoch unvollständig, wenn die gesamtwirtschaftlichen Kosten der Statistikerstellung und der tatsächliche Nutzen der Daten nicht ausreichend berücksichtigt werden. Es muss also die Frage gestellt werden, was ist eine Statistik bzw. eine gewisse Qualität der Statistik dem gesamten System von Wirtschaft und Gesellschaft wirklich wert?

Besonders internationale Organisationen, die sowohl als Nutzer der

Daten als auch als Verkünder der Standards auftreten, scheinen oft zu weit von der Ebene des Melders entfernt zu sein, um diese Zusammenhänge ausreichend würdigen zu können. Aber genau hier liegt der eigentliche Ansatz für gute Qualität in der Statistik: Nur wenn die Melder als Liefe-

ranten der Ausgangsdaten, die Statistikproduzenten und die Nutzer der Statistik als gleichwertige Partner in diesem Prozess der Statistikerzeugung gesehen werden, kann man ausreichende statistische Qualität am Ende dieses Prozesses erwarten.

Literaturverzeichnis

- Bank von Japan. 2001.** Data Quality Assessment Framework. Prepared for the 14th Meeting of the IMF Committee on Balance of Payments Statistics.
- EZB. 2005.** Annual Quality Report on the Euro Area Balance of Payments and International Investment Position Statistics.
- IWF. 2003.** Data Quality Assessment Framework for Balance of Payments Statistics.
- IWF. 2003.** Fifth Review of the Fund's Data Standards Initiatives – Data Quality Assessment Framework and Data Quality Program.
- Joint ECB-Eurostat Task Force on Quality. 2004.** Report on the Quality Assessment of Balance of Payments and International Investment Position.
- Laliberté, L., W. Grünewald, L. Probst. 2004.** Data Quality: A Comparison of IMFs Data Quality Assessment Framework (DQAF) and Eurostat's Quality Definition.
- Office for National Statistics (UK). 2001.** A Framework for Assessing the Quality of Balance of Payments Statistics: An Initial Response by the United Kingdom. Prepared for the 14th meeting of the IMF Committee on Balance of Payments Statistics.