

TEXTE

37/2019

# Quantitative Bewertung von Umweltindikatoren

Sachverständigengutachten



TEXTE 37/2019

Projektnummer 83655

FB000073

# **Quantitative Bewertung von Umweltindikatoren**

Sachverständigengutachten

von

Mark Meyer


Gesellschaft für Wirtschaftliche Strukturforschung mbH,  
Osnabrück


Im Auftrag des Umweltbundesamtes

## Impressum

### Herausgeber

Umweltbundesamt  
Wörlitzer Platz 1  
06844 Dessau-Roßlau  
Tel: +49 340-2103-0  
Fax: +49 340-2103-2285  
[info@umweltbundesamt.de](mailto:info@umweltbundesamt.de)  
Internet: [www.umweltbundesamt.de](http://www.umweltbundesamt.de)

 [/umweltbundesamt.de](https://www.facebook.com/umweltbundesamt.de)

 [/umweltbundesamt](https://twitter.com/umweltbundesamt)

### Durchführung der Studie:

Gesellschaft für Wirtschaftliche Strukturforschung mbH  
Heinrichstraße 30  
49080 Osnabrück

### Abschlussdatum:

November 2017

### Redaktion:

Fachgebiet I 1.5 Nationale und internationale Umweltberichterstattung  
Christian Hoyer, Stephan Timme

Publikationen als pdf:

<http://www.umweltbundesamt.de/publikationen>

ISSN 1862-4804

Dessau-Roßlau, März 2019

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren und spiegeln nicht notwendigerweise die Meinung der GWS mbH oder des Umweltbundesamtes wider.

### **Kurzbeschreibung: Quantitative Bewertung von Umweltindikatoren**

Im Rahmen des Sachverständigengutachtens „Quantitative Bewertung von Umweltindikatoren“ wurde ein R-Programm entwickelt, mit welchem alternative Bewertungsverfahren zur Beurteilung von Umweltindikatoren angewandt werden können. Diese IT-gestützten Analysen historischer Zeitreihenentwicklungen ermöglichen die Beurteilung zukünftiger Trendentwicklungen auf Basis standardisierter statistischer Maßzahlen sowie automatisierter Trend-Identifikations-Algorithmen. Die R-Anwendung ermöglicht dabei eine Anwendung etablierter Bewertungsroutinen der derzeitigen Umweltberichterstattung sowie eines eigenständigen, im Rahmen dieses Vorhabens entwickelten zeitreihenökonometrischen Spezifikations-Algorithmus.

Bei der Indikatorbewertung können sämtliche anwendbaren Verfahren durch die Nutzerinnen und Nutzer flexibel modifiziert werden. Insbesondere der Vergleich der für unterschiedliche Bewertungsverfahren und -vorgaben identifizierten Trendtendenzen ermöglicht eine objektivierte Beurteilung der Frage, ob aktuelle Indikatorentwicklungen auf das zukünftige Erreichen gegebener Indikator-Zielwerte schließen lassen. Zudem können die Auswirkungen alternativer zukünftiger Trendentwicklungen auf die berechneten Zielerreichungsgrade betrachtet werden. Zum Projektabschluss wurde der Quellcode dieser Anwendung dem Umweltbundesamt vollständig übergeben. Das Programm kann daher zukünftig auf beliebigen Arbeitsplatz-PCs des Umweltbundesamtes ohne größeren Aufwand installiert werden.

### **Abstract: Quantitative evaluation of environmental indicators**

This survey documents the implementation of two time series based model selection procedures in R. Both routines feature comprehensive trend identification algorithms which can be automatically applied to empirical datasets. The first routine effectively represents a state of the art trend identification algorithm currently applied within the periodic reporting activities of the German Strategy for Adaptation to Climate Change (“Deutsche Anpassungsstrategie an den Klimawandel”). The second routine introduces a complementary approach which has been newly-created within this project. Both routines have been merged into a single R-application which can be easily executed on usual desktop computers. Moreover, this application also enables users to monitor usual descriptive statistics (like, e.g., rank correlation coefficients or basic trend regression results) and to apply well-established projection methods. These software features facilitate comprehensive data-driven trend assessments in a rather automated way.

## Inhaltsverzeichnis

Abbildungsverzeichnis.....	8
Tabellenverzeichnis.....	9
Abkürzungsverzeichnis.....	10
Zusammenfassung.....	11
Summary .....	20
1 Hintergrund und Aufgabenstellung.....	28
2 Konzeptionelle Grundlagen.....	33
2.1 Wahl der technischen Basis .....	33
2.2 Wahl der zu implementierenden statistischen Methoden.....	34
2.2.1 Deskriptive Statistiken .....	35
2.2.1.1 Zusammenfassende Stichprobeninformationen .....	35
2.2.1.2 Rangkorrelationskoeffizient nach Spearman.....	36
2.2.1.3 Trendidentifikation und -extrapolation gemäß Methode des Statistischen Bundesamtes .....	38
2.2.1.4 Trendidentifikation und -extrapolation gemäß Eurostat .....	39
2.2.2 Explorative Datenanalysen: Trendschätzungsverfahrens „DAS Monitoring“ .....	40
2.2.2.1 Vorstellung des implementierten Algorithmus .....	40
2.2.2.2 Ergänzende Anmerkungen.....	42
2.3 Beurteilung zukünftig zu erwartender Zielerreichungsgrade .....	43
3 Robuste Projektionen zukünftiger Trendverläufe: Identifizierte methodische Ansatzpunkte .....	46
3.1 Theoretische Anmerkungen.....	46
3.1.1 Trendextrapolationen in Anlehnung an Eurostat .....	46
3.1.2 Trendextrapolationen in Anlehnung an DAS Monitoring Verfahren.....	48
3.1.2.1 Herausforderungen bei der Fortschreibung zeitvarianter Trendeffekte.....	48
3.1.2.2 Herausforderungen bei der Identifikation der Ursachen von Autokorrelation. ....	51
3.1.2.3 Beachtung des Scheinkorrelationsrisikos .....	58
3.2 Methodische Weiterentwicklungen im Zuge der Toolentwicklung.....	61
3.2.1 Test des Integrationsgrades.....	61
3.2.1.1 Theoretische Anmerkungen: Herleitung eines Einheitswurzeltests im Kontext der DAS Monitoring Routinen.....	62
3.2.1.2 Im Tool implementierter Testansatz .....	63
3.2.2 Trendprojektionen unter Berücksichtigung des Integrationsgrades der Indikatoren .....	64

3.2.3	Fehler-Korrektur Ansätze als alternativer Ansatz zur Berücksichtigung autokorrelierter Störterme .....	66
4	Abschließende Anmerkungen und Ausblick .....	67
5	Quellenverzeichnis .....	72

## Abbildungsverzeichnis

Abbildung 1:	Exemplarische Ansicht einer Benutzerumgebung bei Anwendung des Analysetools.....	12
Abbildung 2:	Umweltindikator-Berichterstattung auf nationaler und internationaler Ebene (Beispiele).....	13
Abbildung 3:	Grundstrukturen des Schätzalgorithmus von Martone (2013)	15
Abbildung 4:	Abschätzung des Scheinkorrelationsrisikos.....	17
Abbildung 5:	Beispielhafte Darstellung des Berichtsumfangs der Report-Dateien (Auszug).....	35
Abbildung 6:	Beispielhafte Darstellung des Berichtsumfangs der Ergebnisübersicht (Auszug) .....	36
Abbildung 7:	Rangkorrelationskoeffizienten für alternative Trendverläufe	.37
Abbildung 8:	Beispiel einer Trendextrapolation gemäß Methode des Statistischen Bundesamtes.....	38
Abbildung 9:	Beispiel einer Trendextrapolation gemäß Eurostat-Methode	.39
Abbildung 10:	Durch das Trendschätzungsverfahrens „DAS Monitoring“ geprüfte Trendmuster .....	41
Abbildung 11:	Grundstrukturen des Schätzalgorithmus von Martone (2013)	42
Abbildung 12:	Beispiel projizierter Zielerreichungsgrade.....	44
Abbildung 13:	Bewertung von Zielerreichungsgraden, Beispiel der Nachhaltigkeitsberichterstattung des Statistischen Bundesamtes.....	45
Abbildung 14:	Beispiel einer Trendextrapolation in Anlehnung an die Eurostat-Methode .....	46
Abbildung 15:	Beispiel alternativer numerischer Berechnungen von Wachstumsraten .....	47
Abbildung 16:	Beispielhafte Extrapolation eines geschätzten Trendpolynoms.....	49
Abbildung 17:	Empirisches Beispiel zur Frage der Behandlung von Strukturbrüchen am Beispiel Primärenergieverbrauch (in Petajoule) .....	53
Abbildung 18:	Alternative Trendextrapolationen des Primärenergieverbrauchs (in Petajoule) .....	55
Abbildung 19:	Trendextrapolationen unter Verwendung alternativer dynamischer Modellspezifikationen .....	58
Abbildung 20:	Stochastische Trends als Realisation eines simulierten Random Walks .....	60
Abbildung 21:	Abschätzung des Scheinkorrelationsrisikos.....	61
Abbildung 22:	Exemplarische Ansicht einer Benutzerumgebung bei Anwendung des Analysetools.....	67



## Tabellenverzeichnis

Tabelle 1:	Überblick zur (Kern-)Funktionalität der implementierten Bewertungsverfahren.....	31
Tabelle 2:	Entscheidungsregeln bei Anwendung der ADF-Tests.....	64

## Abkürzungsverzeichnis

<b>ADF-Test</b>	Augmented Dickey Fuller Test
<b>AK UGRdL</b>	Arbeitskreis Umweltökonomische Gesamtrechnung der Länder
<b>BMUB</b>	Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit
<b>bspw.</b>	beispielsweise
<b>DAS</b>	Deutsche Anpassungsstrategie an den Klimawandel
<b>Eurostat</b>	Statistisches Amt der Europäischen Union
<b>ggf.</b>	gegebenenfalls
<b>KIS</b>	Umwelt-Kernindikatorensystem
<b>SDGs</b>	Sustainable Development Goals
<b>Stat. Ba</b>	Statistisches Bundesamt
<b>UBA</b>	Umweltbundesamt
<b>UN</b>	Vereinte Nationen

## Zusammenfassung

### Hintergrund

Quantitative Indikatoren sind für eine angemessene Gestaltung von (Umwelt-)Politik sowie zur Überprüfung der Wirksamkeit ausgewählter politischer Maßnahmen unerlässlich. Insbesondere im Bereich der Nachhaltigkeitsforschung unterliegen politisch relevante Entwicklungen üblicherweise multi-dimensionalen Wirkungszusammenhängen, welche aufgrund ihrer Komplexität nicht unmittelbar beobachtet werden können. Quantitative Indikatoren bieten in diesen Fällen eine Möglichkeit zur standardisierten Berichterstattung über aktuelle Entwicklungstendenzen. Die (möglichst regelmäßige) zuverlässige Erhebung entsprechender Indikatoren sowie deren (möglichst zeitnahe) öffentliche Publikation repräsentieren damit ein notwendiges Element zur verlässlichen Beurteilung der aktuellen Situation.

Vor dem Hintergrund dieser Vorbemerkungen ist leicht nachvollziehbar, dass in diversen Politikfeldern umfassende Indikatorberichterstattungen etabliert wurden. Ein (ausgesprochen ambitioniertes) Beispiel für entsprechende Monitoring-Prozesse ist dabei bspw. im Bereich der Nachhaltigkeitspolitik durch die im Jahr 2015 von der UN-Generalversammlung verabschiedeten sogenannten „Sustainable Development Goals“ (SDGs) gegeben.<sup>1</sup> Die SDGs initiieren umfassende Indikatorerhebungen der UN-Mitgliedsstaaten und prägen in der durch sie festgelegten Struktur auch die gegenwärtige Ergebnispräsentation im Indikatorenbericht zur Deutschen Nachhaltigkeitsstrategie (Statistisches Bundesamt 2017).

Neben einer zeitnahen Bestandsaufnahme der aktuellen Situation sind Politik sowie sonstige Stakeholder in der Regel auch an möglichst klaren Aussagen darüber interessiert, ob sich ein Indikator in die gewünschte Richtung entwickelt und ob die Fortsetzung der vergangenen Entwicklung auf die Erreichung der angestrebten Ziele schließen lässt. Werden dabei für die jeweiligen Indikatoren auch konkrete Zielwerte quantifiziert, können politische Maßnahmen in standardisierten Monitoring-Prozessen anhand statistisch beobachtbarer Zielerreichungsgrade evaluiert werden.<sup>2</sup> Dabei können grundsätzlich unterschiedliche Bewertungsansätze verwendet werden. Ein institutions-übergreifend standardisiertes Bewertungsverfahren zur Trendberichterstattung von Nachhaltigkeitsindikatoren wurde bislang nicht etabliert.

---

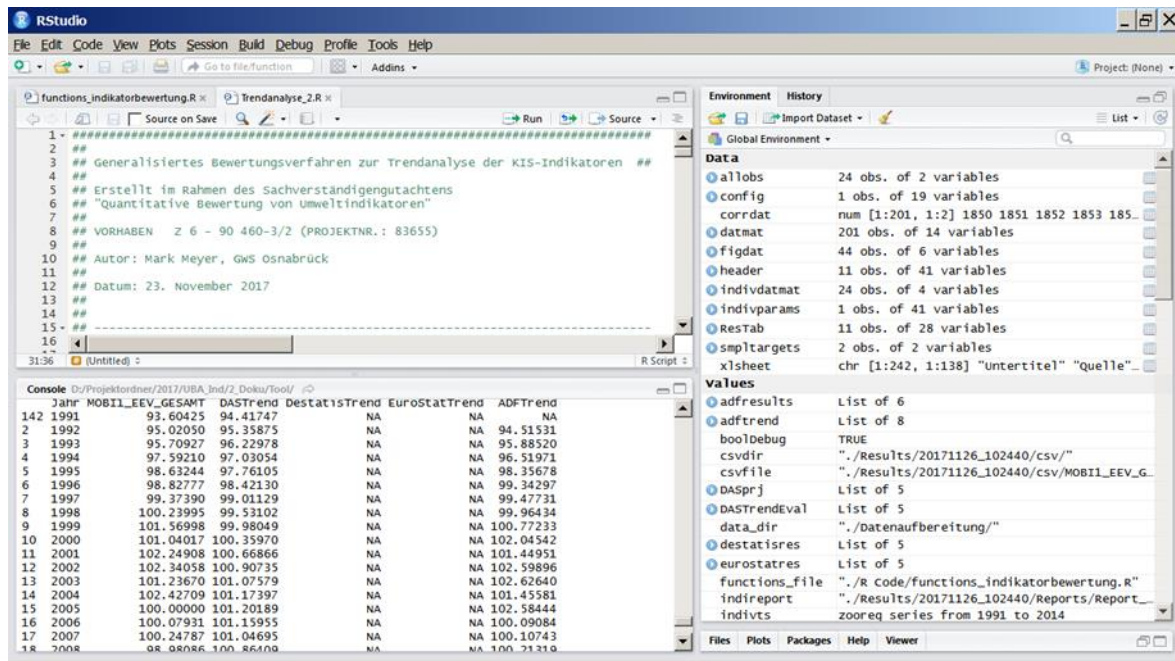
<sup>1</sup> Mit ihrer Agenda 2030 für nachhaltige Entwicklung haben die Vereinten Nationen insgesamt 17 internationale Entwicklungsziele definiert, welche auf eine nachhaltige Verbesserung der Lebensverhältnisse unter Einhaltung globaler Belastungsgrenzen ausgerichtet sind. Zur Messung der Zielerreichung der SDGs wurde dabei eine Liste von 232 globalen Indikatoren entwickelt, die auf Basis nationaler Daten zu befüllen ist. Weiterführende Informationen hierzu sowie zu übrigen Aktivitäten im Bereich der Nachhaltigkeitsberichterstattung finden sich bspw. auch auf der entsprechenden Themenseite des Statistischen Bundesamtes (auf deren Inhalt auch die hier erfolgte Darstellung der SDG-Berichterstattung basiert) unter:

[www.destatis.de/DE/ZahlenFakten/Indikatoren/Nachhaltigkeitsindikatoren/Nachhaltigkeitsindikatoren.html](http://www.destatis.de/DE/ZahlenFakten/Indikatoren/Nachhaltigkeitsindikatoren/Nachhaltigkeitsindikatoren.html)

<sup>2</sup> Um eine möglichst vollständige Darstellung des thematischen Hintergrunds dieses Vorhabens bieten zu können, sei an dieser Stelle kurz ergänzend angemerkt, dass nicht sämtliche verfügbaren Trendprojektionen auf einer empirischen Anwendung formal statistischer Methoden basieren. Neben entsprechenden quantitativen Analysen finden sich in der Literatur auch zahllose qualitative Studien, welche aktuelle Tendenzschätzungen ausgewählter Experten in einer narrativen Zukunftsprojektion zusammenführen. Im Interesse einer möglichst vorausschauenden Politikgestaltung haben solche Ansätze sicherlich ihre Berechtigung. Sie erlauben eine frühzeitige Identifikation von aufziehenden politischen Herausforderungen wie auch Gestaltungsmöglichkeiten ohne dabei durch eine notwendige Verfügbarkeit verlässlicher Indikatoren restringiert zu werden. Siehe bspw. Langsdorf und Hirschnitz-Garbers (2014) für ein aufwendiges Beispiel aus dem Anwendungsgebiet der Ressourcenschonungspolitik.

Poppers positivem Wissenschaftsbegriff folgend (Popper 1994), können diese qualitativen Ansätze in erster Linie als ein Instrument zur (möglichst konsistenten) Theoriebildung interpretiert werden. Die so gewonnenen theoretischen Aussagen sind dann aber, bei Vorliegen valider statistischer Beobachtungen, angemessenen empirischen Falsifikationstests zu unterziehen.

Abbildung 1: Exemplarische Ansicht einer Benutzerumgebung bei Anwendung des Analysetools



Quelle: Eigene Darstellung auf Basis einer exemplarischen Anwendung der R-Anwendung.

Vor einer entsprechenden methodischen Festlegung sind notwendigerweise die durch unterschiedliche Bewertungsmethoden implizierten Trendaussagen umfassend zu evaluieren. Aus diesem Anlass wurde im Rahmen des vorliegenden Gutachtens eine IT-gestützte Anwendung entwickelt, mit welcher entsprechende empirische Analysen weitestgehend automatisiert durchgeführt werden können. Damit konnte dieses Vorhaben eine wesentliche Grundlage für die zukünftig angestrebte Definition einer standardisierten Methodik der Indikatorbewertung am Umweltbundesamt schaffen.

Der Quellcode dieser Anwendung wurde dem Umweltbundesamt zum Projektabschluss gemeinsam mit einem eigenständigen Handbuch zur Dokumentation der Funktionalität der Anwendung sowie relevanter technischer Parameter zur Konfiguration alternativer Funktionalitäten vollständig übergeben. Da die Anwendung in R programmiert wurde (siehe Abbildung 1 für einen beispielhaften Screenshot), kann das Programm auf beliebigen Arbeitsplatz-PCs des Umweltbundesamtes ohne größeren Aufwand zur lizenzfreien Nutzung installiert werden.

Der vorliegende Abschlussbericht dient der nachvollziehbaren Dokumentation sowie kritischen Würdigung der im Tool implementierten Bewertungsroutinen. Neben einer Vorstellung der entsprechenden Tool-Funktionalitäten werden dabei insbesondere methodische Herausforderungen bei der Herleitung robuster Trendprojektionen thematisiert. Wie die folgenden Anmerkungen dieser Zusammenfassung verdeutlichen, werden derzeit von unterschiedlichen Institutionen verschiedene Verfahren zur statistischen Beschreibung historischer Indikatorentwicklungen angewandt. Während das Tool eine flexible Anwendung dieser historischen Analyseverfahren ermöglicht, bleibt es zukünftigen Forschungsaktivitäten vorbehalten, einen Bewertungsalgorithmus zur robusten Projektion zukünftiger Indikatorentwicklungen unter Anwendung entsprechender Analyseansätze zu entwickeln.

## Entwicklung eines eigenständigen R-Tools zur statistischen Trendanalyse sowie zur automatisierten Trendfortschreibung

### Zusammenführung etablierter Bewertungsmethoden

**Abbildung 2: Umweltindikator-Berichterstattung auf nationaler und internationaler Ebene (Beispiele)**



Quellen: Europäische Union (2015), Statistisches Bundesamt (2017), Umweltbundesamt (2017).

Dieses Vorhaben zielte darauf ab, ein IT-gestütztes Tool zur weitestgehend automatisierten Anwendung etablierter Bewertungsverfahren zu erstellen. Die zu diesem Zwecke im Tool implementierten Verfahren sollen an dieser Stelle kurz vorgestellt werden. Exemplarisch wird hierzu zunächst auf die Nachhaltigkeitsberichterstattung des Statistischen Amtes der Europäischen Union (Eurostat) verwiesen (Abbildung 2, links). Für einen gegebenen Indikator werden entsprechende Trendprojektionen von Eurostat auf Basis des geometrischen Mittels des historisch zwischen zwei Referenzjahren berechenbaren Wachstumsfaktors durchgeführt. Daneben führt Abbildung 2 den Indikatorenbericht zur Deutschen Nachhaltigkeitsstrategie auf. Die Status-Übersichten zu einzelnen Indikatoren dieses Berichts basieren ebenfalls auf Trendprojektionen. Dabei wird üblicherweise eine durch die jüngste Beobachtung sowie den Beobachtungswert sechs Jahre zuvor definierte Geradengleichung auf das Zieljahr linear extrapoliert. Der rechte Bereich der Abbildung 2 verweist auf das vom Umweltbundesamt entwickelt „Umwelt-Kernindikatorensystem“ (KIS). Das KIS umfasst derzeit etwa 50 zentrale Umweltindikatoren, deren Entwicklung zukünftig auf Basis einer standardisierten Methode bewertet werden soll.

Als Ausgangspunkt für entsprechende methodische Weiterentwicklungen wurde im Rahmen dieses Vorhabens auch das Trendschätzverfahrens „DAS Monitoring“ im Tool implementiert. Dieser Schätz- und Trendidentifikations-Algorithmus basiert auf umfassenden methodischen Vorarbeiten von Martone (2013) für den Arbeitskreis Umweltökonomische Gesamtrechnungen der Länder (UGRdL). Schönthaler und von Andrian-Werburg (2015) dokumentieren, wie der ursprüngliche Ansatz von Martone in Zusammenarbeit mit dem Statistischen Beratungslabor der Ludwig-Maximilians-Universität (LMU) zur Verstetigung der indikatorbasierten Berichterstattung im Kontext der Deutschen Anpassungsstrategie an den

Klimawandel (DAS) weiterentwickelt wurde.<sup>3</sup> Da die DAS-Monitoring Routinen dem UBA als lauffähige R-Skripte vorlagen, konnte der von Schönthaler und von Andrian-Werburg (2015) dokumentierten DAS-Algorithmus ebenfalls vollständig im Tool implementiert werden.<sup>4</sup> Das Tool ist damit in der Lage, für jeden Indikator die Anpassungsgüte deterministischer Trendregressionen automatisch zu evaluieren und auf Basis dieser Prüfungen die dynamischen Eigenschaften der analysierten Indikatoren auf Basis alternativ getesteter Trendpolynome zu charakterisieren.

Im Vergleich zu sonstigen umweltpolitischen Indikatoranalysen zeichnet sich der Bewertungs-Algorithmus nach Martone (2013) durch einen umfassenden Einsatz zeitreihenanalytischer Methoden aus.<sup>5</sup> Eine vollständige Übersicht der dabei angewandten Modell-Selektionsschritte kann Abbildung 3 entnommen werden.

Wie aus dieser Abbildung ersichtlich wird, kennzeichnet insbesondere die explizite mathematische Modellierung von (signifikant geschätzten) Autokorrelations-Symptomen diesen Algorithmus. Dieses Merkmal reflektiert den Befund, dass bei Anwendung zeitreihenanalytischer Verfahren häufig autokorrelierte Residuen identifiziert werden. Um numerische Verzerrungen der geschätzten Parameter zu vermeiden, wird bei Vorliegen signifikanter Autokorrelationseffekte, in der Tradition von Cochran und Orcutt (1949), eine explizite Schätzung der Autokorrelationsstruktur der Residuen durchgeführt. Zur numerischen Identifikation der jeweiligen Parameter werden dabei iterative Maximum-Likelihood-Schätzer angewandt.

Im Anwendungsfeld der (umwelt-)politischen Indikatoranalyse ist dem Autor kein vergleichbar ausgereifter Algorithmus zur automatisierten Trendanalyse bekannt. Dennoch identifiziert das vorliegende Sachverständigengutachten zwei wesentliche Aspekte, deren methodische Implikationen gegen eine unmittelbare Verwendung des Algorithmus zur Abschätzung zukünftiger Trendentwicklungen sprechen.

Hierzu muss zunächst festgehalten werden, dass die aktuell bei Schönthaler und von Andrian-Werburg (2015) dokumentierte Verwendung dieses Bewertungsverfahrens lediglich eine Bewertung historischer Indikatorentwicklungen anstrebt. Die Schätzung quadrierter Trendterme ist dabei sicherlich gut geeignet, um gegebene Indikatorbeobachtungen auf Anzeichen einer sich im Beobachtungszeitraum verändernden Trenddynamik zu testen. Die signifikante Schätzung eines quadratischen Trendpolynoms weist dann darauf hin, dass die Hypothese eines langfristig stabilen linearen Trends verworfen werden muss. Hieraus darf allerdings nicht gefolgert werden, dass die testweise vorgenommene Approximation des historischen Trendverlaufs auch hinreichend robuste Projektionseigenschaften aufweist. Vielmehr muss zur Kenntnis genommen werden, dass aus zahllosen empirischen Analysen bekannt ist, dass jeder gegebene Zeitreihenverlauf durch Polynome höherer Ordnung mit beliebiger Genauigkeit approximiert werden kann. Für zukünftige Fortschreibungen sind die so approximierten historischen Beobachtungen dann aber häufig ungeeignet.

---

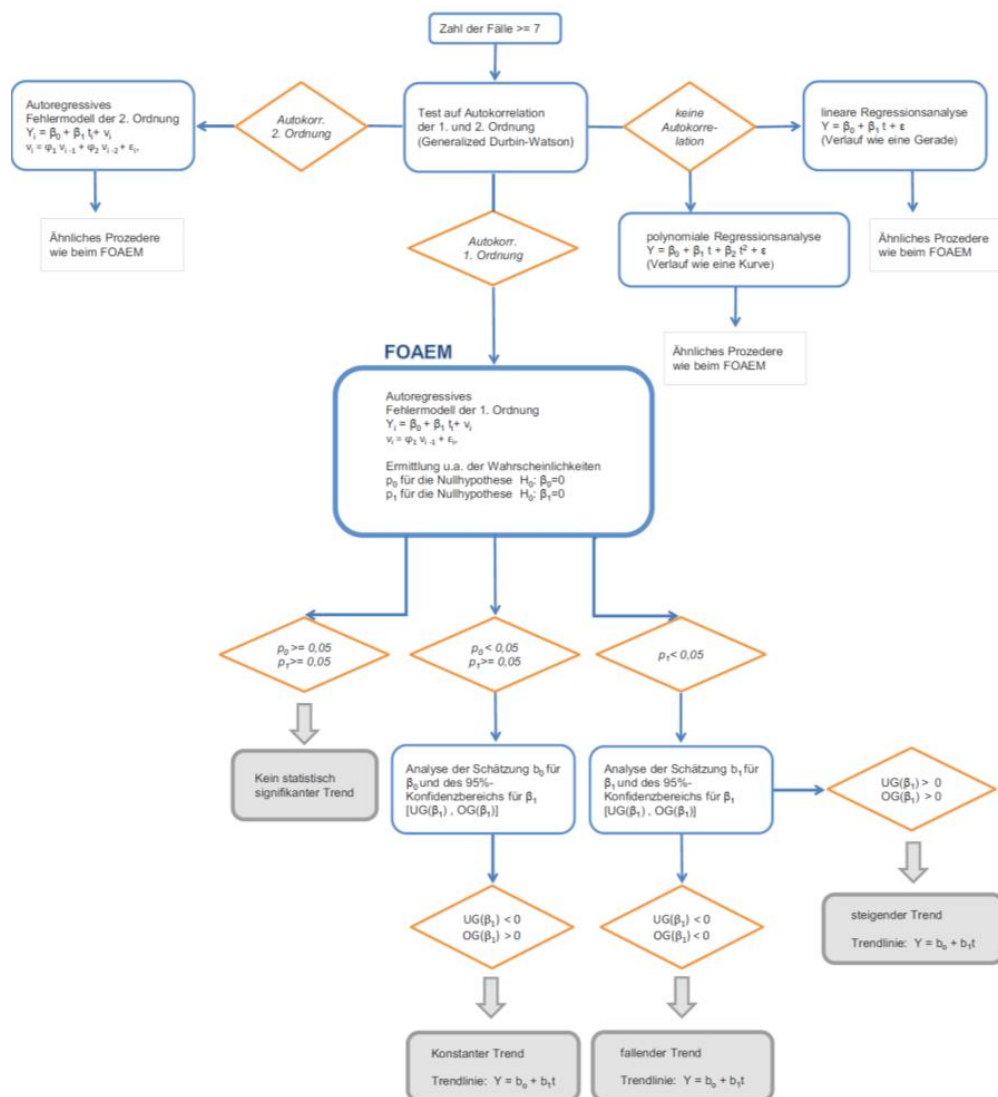
<sup>3</sup> Die Arbeiten der LMU wurden dabei von Prof. Dr. Helmut Küchenhoff verantwortet.

<sup>4</sup> Autor dieses R-Programmcodes (datierend vom Sommer 2014) ist Shuai Shao. Nähere Details der in diesem Code implementierten Funktionalitäten können bei Schönthaler und von Andrian-Werburg (2015) nachgelesen werden.

<sup>5</sup> Die implementierten DAS Monitoring Routinen orientieren sich umfassend an den hier dargestellten Prozessabläufen. Allerdings wird dabei eine Auswahl höhergradiger Trendpolynome tendenziell befördert, indem „... bei Datenreihen mit mehr als 12 Datenpunkten grundsätzlich zunächst quadratische Trends betrachtet werden.“ (Schönthaler und von Andrian-Werburg 2015, S. 53).



Abbildung 3: Grundstrukturen des Schätzalgorithmus von Martone (2013)



Quelle: Martone (2013, S. 86).

Um den zweiten Aspekt, welcher vor einer Anwendung der DAS Routinen zu Projektionszwecken vertieft methodisch diskutiert werden sollte, anschaulich darzustellen zu können sei zunächst daran erinnert, dass die beobachtbaren Entwicklungen der Umweltindikatoren selbstverständlich nicht durch die Zeit verursacht werden. Die Indikatorbefunde resultieren stets aus dem komplexen Zusammenspiel unterschiedlicher und häufig divergierender kausaler Einflussfaktoren. Da eine statistische Identifikation dieser Kausaleffekte in der Regel zeitaufwendig ist, wird bei Trend-Identifikationsverfahren auf eine explizite Berücksichtigung dieser Kausaleffekte verzichtet. Diese Vorgehensweise ist grundsätzlich nicht zu beanstanden. Allerdings sollte beachtet werden, dass dann auch nicht ausgeschlossen werden kann, dass sämtliche Indikatorbeobachtungen lediglich einen zufälligen Befund repräsentieren. Wenn keine inhaltliche Evaluation der für die zeitliche Entwicklung eines Indikators ursächlichen Kausalitäten durchgeführt wird, empfiehlt sich somit eine zusätzliche Prüfung der Hypothese „sämtliche scheinbaren Trendmuster basieren ausschließlich auf zufälligen Indikатораusprägungen“.

In der bei Schönthaler und von Andrian-Werburg (2015) dokumentierten Umsetzung des Trendschätzungsverfahrens zum Monitoring der Deutschen Anpassungsstrategie an den Klimawandel erfolgt keine Prüfung auf die Existenz solcher sogenannter stochastischer Trendmuster. Diese Vorgehensweise kann dadurch legitimiert werden, dass die DAS-Indikatoren erklärterweise „ausschließlich vergangene und gegenwärtige Entwicklungen abbilden [sollen]. Prognosen und Szenarien sind nicht Gegenstand der Darstellungen.“ (Schönthaler und von Andrian-Werburg (2015, S. 1f). – Solch eine beschreibende Abbildung aktueller Indikatorentwicklungen bedarf zunächst sicherlich keiner Klärung der Frage, ob die dabei im Zeitablauf erkennbaren Trendmuster rein zufällig aufgetreten sind. Solange das angewandte Trendschätzungsverfahren nicht zur Projektion zukünftiger Entwicklungen genutzt werden soll, erscheint eine Klärung dieser Frage zumindest aus methodischer Perspektive irrelevant.

Für die im Rahmen dieses Vorhabens zu erstellende R-Anwendung wurde in der Leistungsbeschreibung allerdings explizit gefordert, dass diese die analysierten Indikatoren nicht nur hinsichtlich ihres bisherigen Verlaufs beschreibt, sondern auch die *Signifikanz* identifizierter Trendverläufe abschätzt sowie auf Basis dieser Abschätzungen eine *Projektion zukünftiger Zielerreichungsgrade* ermöglicht. Damit wurde ein Ambitionsniveau vorgegeben, welches den soeben beschriebenen Anwendungsfall einer methodisch-basierten Abbildung historischer Entwicklungsmuster deutlich übersteigt.

#### **Eigenständige Weiterentwicklung des etablierten Instrumentariums zur Indikatorbewertung**

Im Forschungsfeld der ökonometrischen Zeitreihenanalyse ist seit über vierzig Jahren bekannt, dass durch stochastische Trendmuster massive Scheinkorrelationsprobleme hervorgerufen werden können (siehe bspw. Granger und Newbold 1974). Dies bedeutet, dass Signifikanztests unter Verwendung üblicher Verteilungsannahmen im Fall stochastischer Trendmuster unter massiven Verzerrungen in Form einer zu häufig ausgewiesenen Signifikanz leiden können.

Im Verlauf dieses Vorhabens wurde daher ein eigenes Monte-Carlo-Experiment zur quantitativen Abschätzung entsprechender Scheinkorrelations-Risiken bei Anwendung des DAS Monitoring Trendschätzungsverfahrens durchgeführt. Hierfür wurden zunächst 1000 Zeitreihen mit einem Umfang von 200 Beobachtungen als Realisation eines sogenannten „Random Walks“ mit Erwartungswert Null und Standardabweichung fünf simuliert. Für jede so generierte Zeitreihe gilt, dass die beste Prognose durch den letzten beobachteten Wert dieser Zeitreihe gegeben ist.<sup>6</sup> Da die so generierten Zeitreihen keinen deterministischen Trendeinflüssen unterliegen, sollte die Anwendung eines zur Prognose zukünftiger Indikatorentwicklungen genutzten Trendidentifikationsalgorithmus bei diesen Zeitreihen idealerweise zu der Entscheidung tendieren, dass keine langfristige signifikanten Trends identifizierbar sind.

Nachdem jede simulierte Zeitreihe mit dem R-Skript des DAS Monitoring Trendschätzungsverfahrens bewertet wurde, konnten die in Abbildung 4 dargestellten Entscheidungshäufigkeiten beobachtet werden. Damit wird ein erhebliches Scheinkorrelations-Risiko sichtbar: In weniger als 22% aller Fälle entschied das R-Skript richtigerweise, dass kein signifikanter deterministischer Trend vorlag. Dies bedeutet, dass für annähernd vier von fünf Random Walk-Realisationen irrtümlich signifikante Trendanzeichen identifiziert wurden. Formaler ausgedrückt: Wenn eine Zeitreihe stochastische Trendeigenschaften aufweist, kann

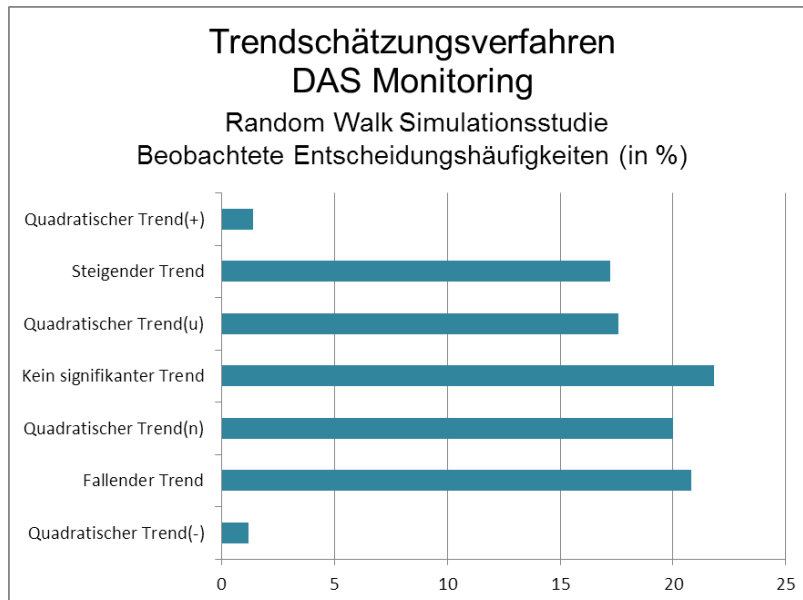
---

<sup>6</sup> In formaler Notation kann der datenerzeugende Prozess des simulierten Random Walks folgendermaßen notiert werden:  $y_t = y_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim iid N(0,25)$ .



das empirische Signifikanzniveau dieser Testroutinen nicht (wie in diesen Simulationen angestrebt) bei 5% sondern bei (annähernd) 80% liegen. Vor diesem Hintergrund ist nach Meinung des Autors von einer unmittelbaren Anwendung des Trendschätzungsverfahrens DAS Monitoring zur Projektion zukünftiger Zielerreichungsgrade abzuraten.

**Abbildung 4: Abschätzung des Scheinkorrelationsrisikos**



Quelle: Eigene Ergebnisse unter Anwendung des in Schönthaler et al. (2015) beschriebenen Schätzalgorithmus.

Dieser Befund verdeutlicht einen akuten Forschungsbedarf zur umfassenden und systematischen Prüfung alternativer methodischer Optionen bei der Fortschreibung gegenwärtiger Indikatorentwicklungen. Der gegebene Projektrahmen zielte allerdings nicht auf entsprechende methodische Weiterentwicklungen ab. Während damit im Rahmen dieses Sachverständigengutachtens keine grundsätzliche Prüfung alternativer methodischer Handlungsoptionen durchgeführt werden konnte, wurde prototypisch eine methodische Option unter Verwendung des sogenannten „Augmented Dickey Fuller Test“ (ADF-Test, vgl. Dickey und Fuller 1979) im Tool implementiert. Diese beispielhafte Umsetzung eines eigenen Trendidentifikations-Algorithmus dient zur Veranschaulichung der Befunde bei Schätzung von differenzierten Indikator-Zeitreihen. Damit kann bei Anwendung des Tools bereits eine Möglichkeit zur Erweiterung bisheriger Bewertungsmethoden durch die ergänzende Einbindung komplementärer Schätz- und Testverfahren exemplarisch angewandt werden. Die systematische Weiterentwicklung der derzeit etablierten Trendbewertungsverfahren unter Berücksichtigung der hierdurch gewonnenen Erkenntnisse sowie unter Verwendung zusätzlicher Testverfahren bleibt allerdings zukünftigen methodischen Forschungsvorhaben vorbehalten.

## Identifizierte Ansatzpunkte zur Weiterentwicklung bisheriger statistischer Trendanalyseverfahren

Aufgrund der soeben angesprochenen verbleibenden (aber im Rahmen dieses Sachverständigengutachtens nicht zu bearbeitenden) wissenschaftlichen Grundfragen einer angemessenen Projektion zeitvariabler Trendeffekte sowie der angemessenen Berücksichtigung von Scheinkorrelationsrisiken und damit offensichtlichen Möglichkeiten zur methodischen Verbesserung seiner derzeit implementierten Bewertungsroutinen, weist die zum Projektabschluss an das Umweltbundesamt übergebene R-Anwendung aktuell einen prototypischen Charakter auf.

Hierzu sei zunächst explizit angemerkt, dass dieses Vorhaben innerhalb der gegebenen Gesamtlaufzeit von weniger als sechs Kalendermonaten aus Sicht des Auftragnehmers ausgesprochen erfolgreich war. Für eventuell in Zukunft erfolgende weitergehende Entwicklungsaktivitäten sollte allerdings eine deutlich längere Bearbeitungszeit erwogen werden, da es sicherlich wünschenswert wäre, sowohl die Grundlagenforschung zu den implementierten Methoden als auch den Zielgruppen-orientierten Ausbau des Tools in einem harmonisierten Vorhaben gemeinsam weiterzuverfolgen.

So war beispielsweise im Rahmen dieses Sachverständigengutachtens keine weitergehende Evaluation des Tools an konkreten Anwendungsbeispielen vorgesehen. Nachdem der im Rahmen dieses Vorhabens erschaffene Prototyp ab sofort sämtlichen Experten des Umweltbundesamtes zur umfassenden Evaluation zur Verfügung steht, könnten zukünftige Entwicklungsaktivitäten darauf abzielen, entsprechende Feedbacks zur Steigerung der Funktionalität der Anwendung einzuholen, um diese bei weiteren Überarbeitungen der Anwendung ebenfalls berücksichtigen zu können.

Dabei sollte auch noch einmal überprüft werden, ob die bislang erfolgten Umsetzungen in R im Interesse einer dauerhaft ausbaufähigen Anwendung in eine alternative Softwareumgebung zu überführen sind. Während die derzeitige Realisation in R im gegebenen Projektrahmen absolut naheliegend war, sollte diese Entscheidung auf keinen Fall bereits als grundsätzliche Festlegung der Softwarebasis für weitergehende Toolentwicklungen verstanden werden. So wurde diese Entscheidung zunächst maßgeblich dadurch beeinflusst, dass für die im Rahmen dieses Vorhabens zu erstellende Anwendung aus Budgetgründen zunächst keine grafische Benutzeroberfläche gestaltet werden sollte. Der zeitliche Entwicklungsaufwand zur Umsetzung der für dieses Sachverständigengutachten benötigten Funktionalität wurde daher nicht wesentlich durch eine Entscheidung zwischen R und alternativen (aber in der Regel lizenzpflichtigen) zeitreihenökonometrischen Softwarepaketen beeinflusst.

Diese Entscheidungssituation kann sich allerdings bei eventuellen späteren Folgeprojekten vollkommen anders darstellen, da in professionellen Anwendungsumgebungen häufig bereits viele Funktionalitäten angelegt sind, welche bei einer Umsetzung in R zunächst (zumindest in einem nennenswerten Umfang) eigenständig programmiert werden müssen. Ohne eine konkrete Beschreibung der benötigten Oberflächen-Eigenschaften lassen sich die bei entsprechenden Weiterentwicklungen in R anfallenden zusätzlichen Entwicklungskosten allerdings kaum verlässlich abschätzen.

Aussagefähiger erscheint stattdessen ein Vergleich mit den zu erwartenden Lizenzkosten bei zukünftiger Verwendung kommerzieller Softwarepakete. Hierzu wird auf das Ökonometriepaket EViews verwiesen, welches Dank umfassender Integration grafischer Bedienelemente eine

intuitive Nutzerführung unter Verwendung interaktiver Anwendungsdialoge ermöglicht.<sup>7</sup> Zum Zeitpunkt der Erstellung dieses Berichts betragen die einmalig notwendigen Lizenzgebühren bei Anschaffung der umfangreichsten EViews-Einzelplatzlizenz weniger als 2000€ (inklusive MwSt).<sup>8</sup> Wenn davon ausgegangen wird, dass maximal zwei Experten simultan Zugang zu diesem Ökonometriepaket benötigen, wären somit insgesamt weniger als 4000€ für die Softwareanschaffung zu budgetieren. Der zusätzlich notwendige Aufwand zur Konzeption und Programmierung einer automatisierten grafischen Benutzerführung, welche die Funktionalität der in diesem Vorhaben erstellten R-Anwendung dann auch in interaktiver Form in EViews zur Verfügung stellt, dürfte das Arbeitsvolumen weniger Personenmonate nicht überschreiten. Dabei ist davon auszugehen, dass sich die EViews-Umsetzung im Vergleich zum bisherigen R-basierten Prototypen insbesondere durch eine schnellere Gesamtperformance, eine übersichtlichere und effizientere Datenorganisation sowie umfassender Flexibilität bei der Implementierung weiterer zukünftiger Funktionalitäten auszeichnen wird.

---

<sup>7</sup> So sind bspw. für in Eviews durchgeführte Schätzungen verschiedene grafische Befunde automatisch per Mausklick abrufbar. Entsprechende Darstellungen können an individuelle Layout-Vorgaben angepasst sowie in gängigen Grafikformaten zur Weiterverarbeitung bzw. Dokumentation exportiert werden (bspw. als \*.eps, \*.jpg, \*.png oder \*.pdf - Dateien). Sämtliche numerischen Ergebnisse können zudem ohne weiteres direkt in Excel-Anwendungen kopiert werden bzw. wiederum in diversen Dateiformaten (bspw. im ASCII-, Excel- oder Lotus-Format) exportiert werden.

<sup>8</sup> Da unterschiedlich umfangreiche Softwarepakete geordert werden können, wird an dieser Stelle die Obergrenze der zusätzlichen Kosten bei Anschaffung der EViews-Software abgeschätzt, indem von einer Anschaffung der teuersten Einzelplatz-Lizenzen zur kommerziellen Nutzung ausgegangen wird. Nähere Details zu den verschiedenen Ausstattungsvarianten können in Deutschland beim Anbieter STATCON erfragt werden (<https://www.statcon.de>). Für öffentliche Forschungseinrichtungen (wie bspw. Universitäten) sind zudem rabattierte Angebote verfügbar.

## Summary

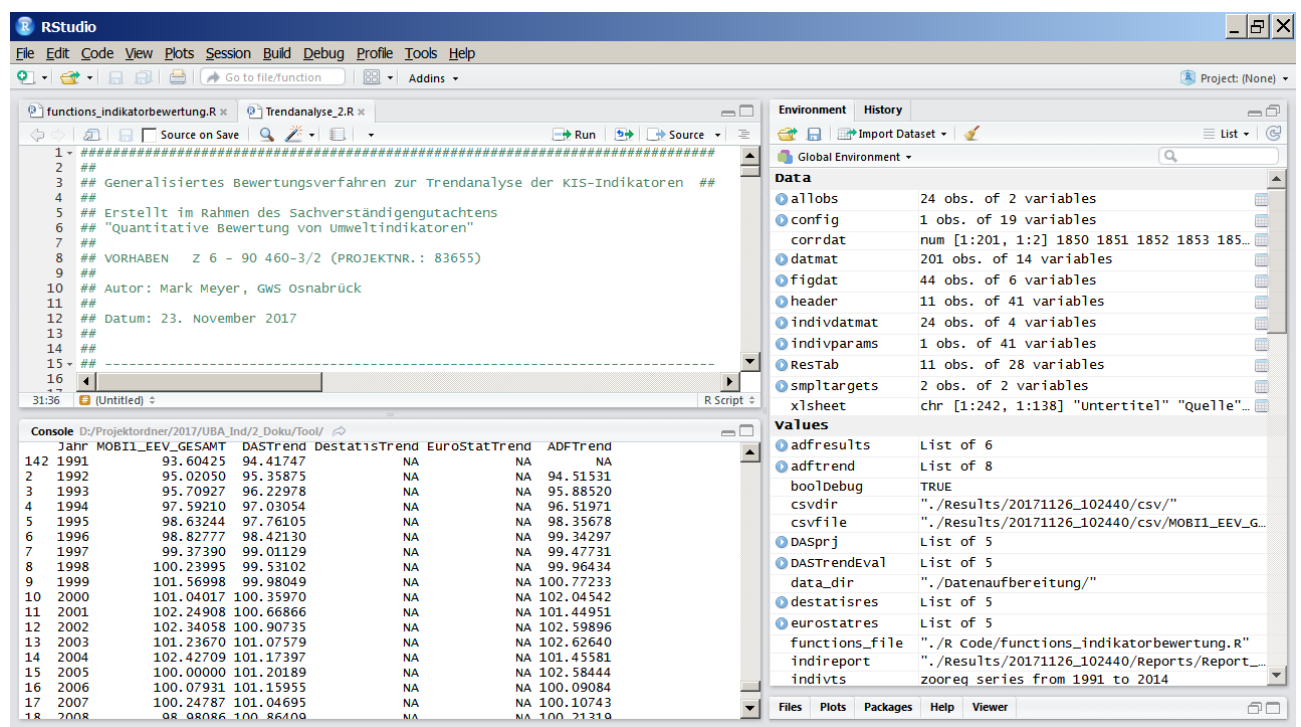
### Introduction

Quantitative indicators are essential for an effective monitoring of applied policy measures. This applies in particular to the area of sustainability research. Sustainability assessments are intended to examine multidimensional developments which tend to be driven by a complex network of drivers and mutual interdependencies. Therefore, within sustainability research underlying causalities of and corresponding individual impacts on political objectives usually cannot be directly observed. Whereas detailed mappings of the driving causal interrelationships usually require comprehensive (and thus time-consuming analyses), quantitative indicators facilitate the in-time reporting about current development trends. The reliable collection of standardized indicators and their periodic publication do therefore represent two necessary elements for any contemporary assessment of current sustainability developments. Quantitative indicators can therefore also be identified as an essential prerequisite for the development of effective (environmental) policy measures.

In coherence with these introductory remarks comprehensive indicator reporting systems have already been established in diverse policy areas. From an environmental viewpoint the Sustainable Development Goals (SDGs) adopted by the UN General Assembly in 2015 provide an (extremely ambitious) example of indicator-based monitoring processes. The SDGs oblige UN member states for comprehensive indicator surveys. In Germany, they are already reflected in the official reports to the German Sustainability Strategy (Statistisches Bundesamt 2017).

Apart from short term indications of current development dynamics, policymakers and other stakeholders are generally also interested in unambiguous information about prospective long term developments. Essentially, this issue refers to the question whether an observed indicators features any significant trend developments. If quantified target values have been defined, standardised monitoring processes may then also derive expected future target achievement rates from applications of statistical estimation routines on historical indicator observations. Methodologically, a variety of individual statistical procedures might be applied for this purpose. A standardised cross-institutional evaluation procedure for trend assessments of sustainability indicators has not yet been established so far. Currently, individual institutions do thus apply different methodologies for similar trend evaluation purposes.

Figure 1: Sample view of the chosen software development environment



Source: Own illustration based on an exemplary application of the software developed within this project.

The present report documents the development of a new software environment which allows for (a largely automated) application of corresponding statistical procedures in own empirical analyses (see Figure 1 for an exemplary screenshot). The source code of this application was handed over to the Federal Environment Agency at the end of the project together with a separate manual documenting the functionality of the application and relevant technical parameters. As the application was programmed in R it can be installed on any workstation PC of the Federal Environment Agency without major efforts for licence-free use. Thereby, this project contributed an essential tool for future research work towards the establishment of a standardised methodology for indicator evaluation at the Federal Environment Agency.

The report provides an overview on the evaluation routines implemented in the tool. In addition to an introduction to corresponding tool functionalities, the presentation focuses on a presentation of fundamental methodological challenges in the derivation of robust trend forecasts. In this regard it is important to note that this project was not intended to foster methodological research. Therefore, the task of developing a self-contained evaluation algorithm for the robust projection of future indicator developments is left for future research activities. According to the author's view, the Federal Environment Agency is well advised to promote in-depth research concerning the methodological challenges emphasised in this report by future project activities.

## Development of a self-contained software for statistical trend analysis and automated trend projections

### Commonly applied trend identification methods

Figure 2: Exemplary references to state of the art environmental indicator reports



Source: Europäische Union (2015), Statistisches Bundesamt (2017), Umweltbundesamt (2017).

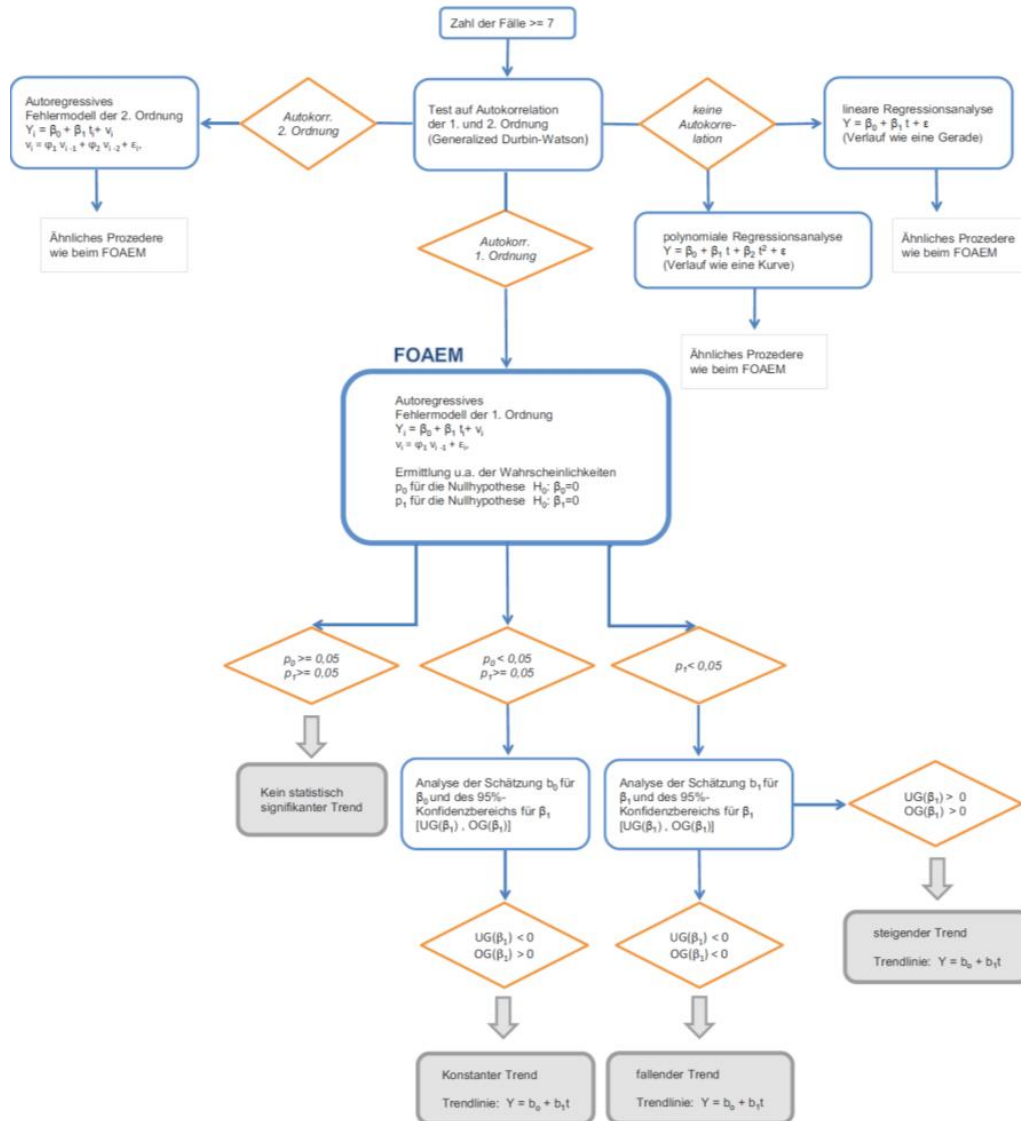
The aim of this project was to create an IT-supported tool for the largely automated application of commonly adopted trend evaluation procedures. The implemented procedures will be briefly presented here: The first example refers to the sustainability reporting of the Statistical Office of the European Union (Eurostat) (Figure 2, left). For a given indicator, Eurostat's trend projections are based on the geometric mean of the growth factor calculated between two reference years. Figure 2 also lists the indicator report for the German sustainability strategy. In this report, status overviews for individual indicators also feature individual trend projections. These projections are determined by an extrapolation of a straight line through the most recent indicator observation and the indicator observation six years before. The right-hand section of Figure 2 refers to the "Environmental Core Indicator System" ("Umwelt-Kernindikatorensystem", KIS) developed by the Federal Environment Agency. KIS currently comprises around 50 central environmental indicators. Until now, the Federal Environment Agency did not determine an original indicator evaluation method for own trend assessments.

The "DAS Monitoring" trend estimation procedure was also implemented in the tool. This estimation and trend identification algorithm is based on seminal methodological work by Martone (2013) for the German Working Group on Environmental-Economic Accounting of Federal States ("Arbeitskreis Umweltökonomische Gesamtrechnungen der Länder", UGRdL). Schönthaler and Andrian-Werburg (2015) document how Martone's original approach has been further developed in cooperation with the Statistical Consulting Laboratory of the Ludwig-Maximilians-University (LMU) towards a harmonised indicator-based monitoring routine for the German Strategy for Adaptation to Climate Change (DAS). These DAS monitoring routines were already available to UBA as executable R scripts. Hence, they have also been fully implemented in the tool. For any given indicator the tool thus facilitates an automatic evaluation of the historical fit for alternatively tested trend polynomials. Compared to other environmental policy



indicator analyses, Martone’s original algorithm already featured comprehensive applications of time series methods. A complete overview of the implied model selection steps can be inferred from Figure 3.

Figure 3: Basic structures of Martone's estimation algorithm (2013)



Source: Martone (2013, p. 86).

This algorithm is notably characterised by its explicit mathematical modeling of (significantly estimated) autocorrelation patterns. This feature refers to the fact that applications of time-series analytical methods often suffer from autocorrelated residuals. Therefore, this algorithm involves automatic tests for autocorrelated residuals. In case of significant indications of autocorrelated residuals, correlation patterns together with underlying trend parameters are estimated by applications of iterative maximum likelihood estimators in the tradition of Cochrane and Orcutt (1949). The author is not aware of any comparably comprehensive algorithm applications for automated trend analyses of (environmental) policy indicators.

Nevertheless, two essential aspects warn against direct applications of this algorithm for future trend projections.

To this it must first be noted that the DAS Monitoring routines documented in Schönthaler and Andrian-Werburg (2015) merely aim at descriptive classifications of historical indicator developments. For that purpose, it is certainly legitimate to estimation the significance of squared trend terms in order to test given indicator observations for unstable trend dynamics. The significant estimate of a quadratic trend polynomial then indicates that the hypothesis of a long-term stable linear trend must be rejected. However, this does not imply that the estimated test regression will also feature robust projection properties: Within a given sample, any indicator series may be approximated with arbitrary precision by higher order trend polynomials. However, out-of-sample, higher order trend polynomials tend to generate inadequate projections.

Second, it should be noticed that any observed indicator dynamic is of course not caused by time. The observable indicator developments always represent overall findings which emerged from a complex interplay of different and often divergent causal factors. Generally, one may also attempt to identify individual causing effects by applications of elaborated multivariate statistical estimation procedures. However, due to an immanent complexity of the respective methods, any application of such structural identification procedures will generally require a high level of econometric expertise. Such studies are well-justified for empirical assessments of mutual interdependencies. In fact, they may be interpreted as a basic prerequisite for any effective design of individual policy measures. Though, they are actually not required for an in-time monitoring of current indicator dynamics and a deduction of implied long term development prospects. Short-term monitoring routines can therefore refrain from detailed mappings of implied structural causalities.

However, in absence of any causal explanations it must be noted that any given indicator series may simply represent arbitrary trend patterns which emerged by mere chance. Hence, in absence of any causal explanations, statistical tests of the "apparent trend patterns evolved by chance" hypothesis appear as mandatory building block of any thorough trend projection procedure.

The DAS Monitoring routines documented in Schönthaler and Andrian-Werburg (2015) have not been developed for trend projection purposes. Therefore, they do not feature any testing routines for the just mentioned (so-called stochastic trend) hypothesis. As a long as these routines are only applied for descriptive illustrations of current indicator developments, this is not to be criticised. However, this issue questions any attempts for direct applications of the DAS Monitoring routines for future trend projections.

For the R application developed within this project the latter case applies as this project aimed to provide a self-contained software tool which, i.a., infers the statistical significance of identified trend processes in order to judge ex ante trend projections in light of quantified target values. This ambition level clearly exceeds usual ambition levels of the DAS Monitoring routines.

### **Own methodological research activities**

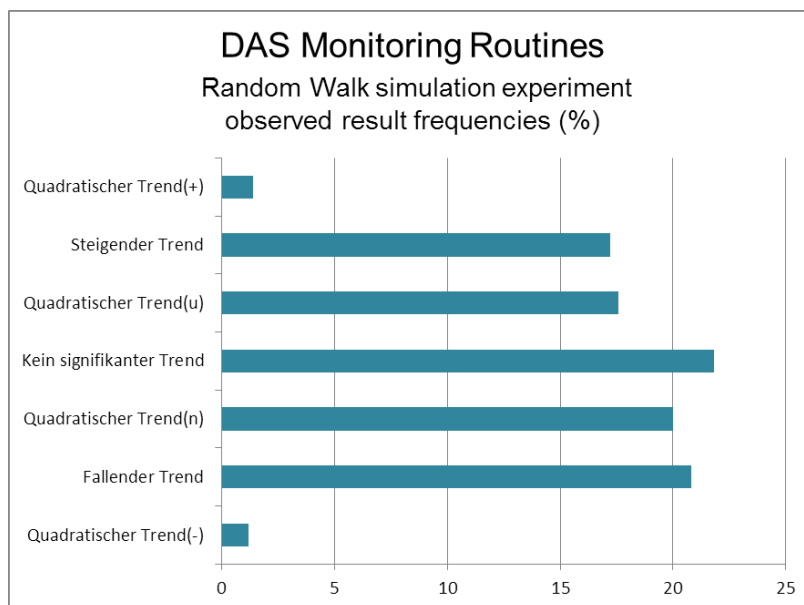
Econometric research already provided ample evidence of the fact that many regression analyses of trending time series may actually suffer from considerable so-called spurious regression effects (see, e.g., Granger and Newbold 1974 for a seminal contribution in this



regard). This means that, if stochastic trend patterns exist, usual interpretations of commonly applied test statistics (like, e.g., the t-statistic) appear too frequently suggestive of significant regression results.

For this project the relevance of such so-called size distortions for the DAS Monitoring routines has been exemplarily demonstrated by means of an own Monte Carlo simulation experiment. To this, 1000 artificial indicator time series, each containing 200 observations, were simulated as realizations of a so-called "random walk" process with an expected value of zero and a standard deviation of five. For these simulated random walk processes, the best forecast of future developments would always be given by the last available historical observation. In other words: None of the simulated indicator series is subject to any deterministic trend influences which could be reasonably projected into the future. Accordingly, any application of a trend identification algorithm intended to forecast future indicator developments should ideally tend to the decision that for these series no significant trends can be statistically inferred.

**Figure 4: Implied spurious regression risks of applications of the DAS Monitoring routines on integrated time series**



Source: Own results from an original Monte Carlo experiment based on the estimation algorithm described in Schönthaler et al. (2015).

Each simulated time series was then evaluated by an application of the R-script of the DAS Monitoring procedure. As illustrated in Figure 4, the R-script decided only in less than 22% of all cases that the analysed series featured no significant deterministic trend. This means that spurious regression results affected the reported findings in nearly four out of five analyses. More formally: In this simulation setup, which applied usual critical values of a 5% significance level, the empirical significance level of the test routines equalled almost 80%. These findings warn against any direct applications of the DAS Monitoring procedure in ex ante trend projection studies.

This finding illustrates an acute need for research into the comprehensive and systematic examination of alternative methodological options in the updating of current indicator developments. The given project framework, however, did not aim at corresponding further

methodological developments. While it was not possible to carry out a fundamental examination of alternative methodological options for action within the framework of this expert report, a methodological option using the so-called "Augmented Dickey Fuller Test" (ADF test, cf. Dickey and Fuller 1979) was implemented in the tool. This exemplary implementation of an own trend identification algorithm serves to illustrate the findings by estimating differentiated indicator time series. Thus, when using the tool, a possibility to extend existing evaluation methods can already be applied exemplarily by the supplementary integration of complementary estimation and test procedures. However, the systematic further development of the currently established trend evaluation methods, taking into account the knowledge gained in this way and using additional test methods, remains reserved for future mathematical research projects.

### **Outlook on prospective further developments**

The given project framework was not intended to promote new methodological developments. Therefore, this report is neither intended to deepen respective methodological discussions nor to provide a comprehensive review of possible amendments and modifications of the DAS Monitoring procedures which might effectively curtail the spurious regression risk. Nevertheless, the just-mentioned spurious regression risk definitely asks for further intensified future methodological refinements of currently applied trend identification algorithms. Hence, in order to provide a selected complementary identification approach, the R-tool is distinguished by the demonstrative implementation of an own trend identification algorithm based on so-called "Augmented Dickey Fuller Test" (ADF test, Dickey and Fuller 1979). This exemplary own trend identification algorithm illustrates the applicability of usual statistical testing routines to differentiated time series. This development work enables the Federal Environment Agency to judge the sensitivity of individual trend projections to selected methodological issues.

However, a systemic assessment of the broader statistical implications of different projection approaches remains for future research. According to the author's view, the Federal Environment Agency would be well advised to foster such activities in supplemental future research projects. The same applies for further key research questions which have been identified within this project but which could not be intensively discussed within this report. Notably the treatment of time-variant trend effects might be approached by a variety of statistical methods which, however, could not be further examined within the given project scope. Overall, the R-tool handed over to the Federal Environment Agency at the end of this project does therefore feature some prototypical characteristics which may certainly be further refined by supplemental research.

Notwithstanding all remaining methodological refinement options, this project was definitely extremely successful. Whereas all achievements documented in this report have been delivered in less than six calendar months, eventual supplemental research activities might benefit from prolonged project durations. Notably, as it would certainly be desirable to consolidate both, basic research on implemented methods as well target-group-oriented tool refinements, together in a harmonized project. Concerning usability, it is important to note that design and development of individual handling aspects (like, e.g., interactive user guidance) had not been budgeted for this project. Future development activities would therefore certainly also be able to increase the usability of this software (e.g., by a systemic review of individual user feedback from applications of the current prototype).

In light of prospective further tool developments the present decision for a software implementation in R should also be reappraised. Within the given project framework, the implementation in R more or less suggested itself. But it is important to note that this decision was (i.a.) significantly influenced by the fact that no graphical user interface had to be designed for this software prototype. Furthermore, on behalf of the funding agency, the DAS Monitoring routines (which are coded in R) should be implemented within this tool. Therefore, within the given project scope, an R-implementation was identified as cost-effective solution.

But this should not be understood as a seminal definition of the software basis for all prospective further tool developments. Depending on the required methodological complexity as well as aspired usability levels for the refined tool, an exchange of the current software environment may facilitate future development works to considerable degrees. As a matter of fact, professional econometric software packages do provide interactive access to many statistical assessment routines. As far as license fees apply, these functionalities are not provided by free. But, a stand-alone implementation of similar functionalities in R will (at least to a considerable extent) require additional development efforts which may easily exceed respective licence fees.

In absence of a concrete definition of future tool capabilities it is of course hardly possible to evaluate the additional development costs which may incur from further software refinements in R. However, an outlook on the prospective license costs for future applications of commercial software packages reveals rather moderate acquisition costs. For this, reference is made to the econometrics package EViews. Due to its comprehensive integration of graphical control elements, EViews facilitates intuitive user guidance based on interactive application dialogs. When this report was written, the one-time license fees for a purchase of the most comprehensive EViews single-user license amounted to less than 2000€ (including VAT). Therefore, assuming that a maximum of two experts need simultaneous access to this econometrics package, a total of less than 4000€ would have to be budgeted for the software acquisition.

## 1 Hintergrund und Aufgabenstellung

Im Bereich der Nachhaltigkeitsforschung unterliegen politisch relevante Entwicklungen üblicherweise multi-dimensionalen Wirkungszusammenhängen, welche aufgrund ihrer Komplexität nicht unmittelbar beobachtet werden können. Quantitative Indikatoren bieten in diesen Fällen eine Möglichkeit zur standardisierten Berichterstattung über aktuelle Entwicklungstendenzen. Die (möglichst regelmäßige) zuverlässige Erhebung entsprechender Indikatoren sowie deren (möglichst zeitnahe) öffentliche Publikation repräsentieren damit ein notwendiges Element zur verlässlichen Beurteilung der aktuellen Situation.

In diversen Politikfeldern wurden inzwischen umfassende Indikatorberichterstattungen etabliert. Ein (ausgesprochen ambitioniertes) Beispiel für entsprechende Monitoring-Prozesse ist dabei bspw. im Bereich der Nachhaltigkeitspolitik durch die im Jahr 2015 von der UN-Generalversammlung verabschiedeten sogenannten „Sustainable Development Goals“ (SDGs) gegeben.<sup>9</sup> Die SDGs initiieren umfassende Indikatorerhebungen der UN-Mitgliedsstaaten und prägen in der durch sie festgelegten Struktur bspw. auch die gegenwärtige Ergebnispräsentation im Indikatorenbericht zur Deutschen Nachhaltigkeitsstrategie (Statistisches Bundesamt 2017). Ein weiteres Beispiel für vergleichbare Indikator-Erhebungen ist in Deutschland bspw. auch durch entsprechende Monitoring-Verfahren im Zuge der Berichterstattung zur Umsetzung der Deutschen Anpassungsstrategie an den Klimawandel (DAS) gegeben (Schönthaler und von Andrian-Werburg (2015)). Beobachtete Indikatorentwicklungen werden dabei im Rahmen eines systematischen Berichtswesens erfasst und transparent dargestellt. Unter Anwendung statistischer Bewertungsmethoden kann zudem eine objektiv nachvollziehbare Abschätzung erfolgen, in welche Richtung sich die analysierten Indikatoren gegenwärtig entwickeln.

Darüber hinaus sind sowohl die Politik als auch weitere Stakeholder grundsätzlich an möglichst klaren Aussagen dahingehend interessiert, ob bei einer Fortsetzung der gegenwärtigen Entwicklungstendenzen eines betrachteten Indikators auch das Erreichen zukünftiger Zielwerte erwartet werden kann. In der amtlichen Statistik werden diesbezüglich derzeit verschiedene statistische Fortschreibungsroutinen verwendet (Eurostat 2014, Statistisches Bundesamt 2017). Da diese allerdings keine Beurteilung der Signifikanz historischer Trendverläufe anstreben, existiert bislang keine einheitliche Methode zur robusten Projektion zukünftig erwartbarer Indikatorverläufe.<sup>10</sup>

<sup>9</sup> Mit ihrer Agenda 2030 für nachhaltige Entwicklung haben die Vereinten Nationen insgesamt 17 internationale Entwicklungsziele definiert, welche auf eine nachhaltige Verbesserung der Lebensverhältnisse unter Einhaltung globaler Belastungsgrenzen ausgerichtet sind. Zur Messung der Zielerreichung der SDGs wurde dabei eine Liste von 232 globalen Indikatoren entwickelt, die auf Basis nationaler Daten zu befüllen ist. Weiterführende Informationen hierzu sowie zu übrigen Aktivitäten im Bereich der Nachhaltigkeitsberichterstattung finden sich bspw. auch auf der entsprechenden Themenseite des Statistischen Bundesamtes (auf deren Inhalt auch die hier erfolgte Darstellung der SDG-Berichterstattung basiert) unter:

[www.destatis.de/DE/ZahlenFakten/Indikatoren/Nachhaltigkeitsindikatoren/Nachhaltigkeitsindikatoren.html](http://www.destatis.de/DE/ZahlenFakten/Indikatoren/Nachhaltigkeitsindikatoren/Nachhaltigkeitsindikatoren.html).

<sup>10</sup> Um eine möglichst vollständige Darstellung des thematischen Hintergrunds dieses Vorhabens bieten zu können, sei an dieser Stelle kurz ergänzend angemerkt, dass nicht sämtliche verfügbaren Trendprojektionen auf einer empirischen Anwendung formal statistischer Methoden basieren. Neben entsprechenden quantitativen Analysen finden sich in der Literatur auch zahllose qualitative Studien, welche aktuelle Tendenzeinschätzungen ausgewählter Experten in einer narrativen Zukunftsprojektion zusammenführen. Im Interesse einer möglichst vorausschauenden Politikgestaltung haben solche Ansätze sicherlich ihre Berechtigung. Sie erlauben eine frühzeitige Identifikation von aufziehenden politischen Herausforderungen wie auch Gestaltungsmöglichkeiten ohne dabei durch eine notwendige Verfügbarkeit verlässlicher Indikatoren restringiert zu werden. Siehe bspw. Langsdorf und Hirschnitz-Garbers (2014) für ein aufwendiges Beispiel aus dem Anwendungsgebiet der Ressourcenschonungspolitik. Poppers positivem Wissenschaftsbegriff folgend (Popper 1994), können diese qualitativen Ansätze in erster Linie als ein Instrument zur (möglichst konsistenten) Theoriebildung interpretiert werden. Die so gewonnenen theoretischen Aussagen sind dann

Das Umweltbundesamt hat das „Umwelt-Kernindikatorensystem“ (KIS) entwickelt, welches etwa 50 zentrale Umweltindikatoren umfasst. Während die Etablierung einer einheitlichen Methode zur robusten Projektion dieser Indikatoren ebenfalls vertieften zukünftigen methodischen Forschungsaktivitäten vorbehalten bleiben muss, konnte dem Umweltbundesamt damit bereits ein notwendiges Werkzeug zur Beurteilung der Leistungsfähigkeit und Sensitivität der derzeit in der angewandten Indikatorberichterstattung gängigen Bewertungsroutinen zur Verfügung gestellt werden. Im Rahmen der nachfolgend dokumentierten Arbeiten wurde eine prototypische Softwareanwendung entwickelt, die eine flexible und dabei weitestgehend standardisierte Bewertung von Indikatoren ermöglicht. Wie vom Auftraggeber gefordert, kann damit eine automatisierte Bewertung einer größeren Anzahl von Indikatoren anhand individuell konfigurierbarer Bewertungsparameter durchgeführt werden.

Die wesentlichen Bestandteile der implementierten Bewertungsroutinen wurden durch die Leistungsbeschreibung des Projekts vorgegeben und lassen sich folgendermaßen zusammenfassen:

- ▶ Analyse des historischen Verlaufs der Indikatoren (Identifikation von Trendmustern),
- ▶ Fortschreibung zukünftiger Trendentwicklungen,
- ▶ Prüfung der Stabilität identifizierter Trendmuster (Signifikanz geschätzter Schätzfunktionen),
- ▶ Abschätzung zukünftig erwartbarer Zielerreichungsgrade.

Eine Übersicht der für diese Zwecke in der Softwareanwendung implementierten Bewertungsroutinen kann Tabelle 1 entnommen werden. Sämtliche dort aufgeführten Routinen werden beim Programmablauf grundsätzlich automatisch ausgeführt.

Grundsätzlich ist dabei darauf hinzuweisen, dass die in dieser Tabelle (u.a.) aufgeführten Bewertungsverfahren in Anlehnung an Vorgehensweisen des Statistischen Bundesamts sowie von Eurostat die jeweiligen Bewertungsroutinen dieser Institutionen nicht ausschließlich reproduzieren. Auf Wunsch des Auftraggebers zeichnet sich die Softwareanwendung durch eine umfassende Flexibilität aus. Sämtliche relevanten Bewertungsparameter der jeweiligen Verfahren können dabei von den Nutzerinnen und Nutzern vor Durchführung einer Bewertung individuell spezifiziert werden, wodurch eine einfache Abschätzung der Sensitivität der Bewertungsergebnisse im Hinblick auf übliche Parametervorgaben ermöglicht wird. Im weiteren Verlauf dieses Textes werden daher auch beispielhafte Tool-Ergebnisgrafiken präsentiert, welche bspw. Trendextrapolationen auf Basis der Eurostat-Methode präsentieren, dabei aber nicht exakt die übliche Eurostat-Vorgehensweise reproduzieren.<sup>11</sup>

Mit Projektende wurde dem Umweltbundesamt der umfassend kommentierte Quellcode der entwickelten Anwendung vollständig übergeben. Das Programm kann auf beliebigen Arbeitsplatz-PCs des Umweltbundesamtes ohne größeren Aufwand lizenzfrei zur eigenständigen Nutzung durch die jeweiligen Anwenderinnen und Anwender installiert werden.

---

aber, bei Vorliegen valider statistischer Beobachtungen, angemessenen empirischen Falsifikationstests zu unterziehen.

<sup>11</sup> In der Eurostat-Berichterstattung werden stets feste Zeitintervalle (5 Jahre / 10 Jahre) analysiert (vgl. hierzu bspw. Eurostat 2014). Im Gegensatz hierzu erlaubt die im Rahmen dieses Vorhabens erstellte Softwareanwendung eine freie Wahl des zu analysierenden Zeitraums.

Die nähere Dokumentation der Funktionalität dieser Anwendung sowie relevanter technischer Parameter zur Konfiguration der in Tabelle 1 aufgeführten Bewertungsroutinen erfolgt in einem eigenständigen Benutzerhandbuch. Da das Benutzerhandbuch dem Umweltbundesamt ebenfalls vorliegt, wird an dieser Stelle auf eine weitergehende Diskussion technischer Details bei der Bedienung der Anwendung verzichtet.

**Tabelle 1: Überblick zur (Kern-)Funktionalität der implementierten Bewertungsverfahren**

Analyseschritt	Bewertungsroutine	inhaltliche Kurzbeschreibung der Funktionalität
Identifikation von Trendmustern	Rangkorrelationskoeffizient nach Spearman	Maßzahl zur Beurteilung der Monotonieeigenschaften einer Indikatorzeitreihe.
	Trendextrapolationsverfahren in Anlehnung an die Methode des Statistischen Bundesamtes	Berechnung des linearen Trends einer Indikatorzeitreihe als arithmetisches Mittel jährlicher absoluter Veränderungen eines analysierten Zeitintervalls.
	Trendextrapolationsverfahren in Anlehnung an die Methode von Eurostat	Berechnung der durchschnittlichen jährlichen Wachstumsrate einer Indikatorzeitreihe als geometrisches Mittel des Wachstumsfaktors eines analysierten Zeitintervalls.
	Trendschätzungsverfahren DAS Monitoring	Statistischer Identifikationsalgorithmus, welcher für eine gegebene Indikatorzeitreihe alternative Trendhypothesen testet. (Prüfung der Anpassungsgüte alternativer Trendpolynome für die Niveauewerte des Indikators).
	Weiterentwicklung des bislang etablierten Instrumentariums zur Indikatorbewertung	Statistischer Identifikationsalgorithmus, welcher für eine gegebene Indikatorzeitreihe alternative Trendhypothesen testet. (Prüfung der Anpassungsgüte alternativer Trendpolynome für die differenzierten Niveauewerte des Indikators).
Fortschreibung zukünftiger Trendentwicklungen	Trendextrapolationsverfahren in Anlehnung an die Methode des Statistischen Bundesamtes	Lineare Fortschreibung der Indikatorzeitreihe auf Basis der zuletzt am aktuellen Rand beobachteten durchschnittlichen Veränderungen.
	Trendextrapolationsverfahren in Anlehnung an die Methode von Eurostat	Exponentielle Fortschreibung der Indikatorzeitreihe auf Basis des geometrischen Mittels des am aktuellen Rand beobachteten Wachstumsfaktors
	Trendschätzungsverfahren DAS Monitoring	Gemäß der Befunde des zuvor angewandten Prüfalgorithmus werden die Niveauewerte des Indikators ggf. mit linearen oder quadratischen Trendpolynomen fortgeschrieben.
	Weiterentwicklung des bislang etablierten Instrumentariums zur Indikatorbewertung	Gemäß der Befunde des zuvor angewandten Prüfalgorithmus werden die Veränderungen des Indikators ggf. fortgeschrieben. Für die betrachtete Indikatorzeitreihe können hieraus lineare oder allgemein Trendverläufe resultieren.
Signifikanztest geschätzter Schätzfunktionen	Trendschätzungsverfahren DAS Monitoring	Der Prüfalgorithmus basiert auf der Verwendung alternativer statistischer Schätzverfahren. Für alle geschätzten Trendverläufe können gängige Teststatistiken (bspw. t-Statistik) ausgewiesen werden.
	Weiterentwicklung des bislang etablierten Instrumentariums zur Indikatorbewertung	Der Prüfalgorithmus basiert auf der Verwendung des Schätzverfahrens der Kleinsten Quadrate. Zudem erfolgt ein Test des Integrationsgrades der analysierten Zeitreihe (ADF-Test). Für alle geschätzten Trendverläufe können gängige Teststatistiken (bspw. t-Statistik) ausgewiesen werden.
Abschätzung zukünftig erwartbarer Zielerreichungsgrade	Zielerreichungsgrad gemäß Eurostat	Berechnung der Relation von Wachstumsraten eines Indikators. Die durchschnittlich projizierte Wachstumsrate zwischen der letzten historischen Beobachtung sowie dem vorgegebenen Zieljahr wird in Relation zu der für eine Einhaltung des Zielwertes im Zieljahr durchschnittlich notwendigen Wachstumsrate betrachtet.
	Eigene Metrik zur Abbildung des Zielerreichungsgrads (einzelner Zielwert)	Berechnung der Relation absoluter Veränderungen eines Indikators. Die projizierte absolute Veränderung zwischen der letzten historischen Beobachtung sowie dem vorgegebenen Zieljahr wird in Relation zu der für eine Einhaltung des Zielwertes notwendigen absoluten Veränderung betrachtet.
	Eigene Metrik zur Abbildung des Zielerreichungsgrads für mehrere Zielwerte	Wenn im Projektionszeitraum mehrere Zielwerte betrachtet werden können, wird bei Trendextrapolationen automatisch der prozentuale Anteil der erreichten Zielvorgaben ausgegeben.

Ergänzend dient das vorliegende Sachverständigengutachten einer methodischen Vorstellung relevanter Eigenschaften der in der obigen Tabelle aufgeführten Bewertungsroutinen. Anhand anschaulicher Beispiele wird dabei für eine (im Vergleich zu derzeit gängigen Verfahren) intensiviertere Nutzung induktiv statistischer Analyseansätze in der Indikatorbewertung argumentiert. Diesbezüglich erscheinen mittelfristig weitaus umfangreichere methodische Erweiterungen der derzeitigen Tool-Funktionalitäten möglich. Nachdem im Rahmen dieses Sachverständigengutachtens allerdings keine methodische Grundlagenforschung erfolgen sollte, muss eine umfassende empirische Evaluation ausgewählter statistischen Verfahren zukünftigen Forschungsprojekten vorbehalten bleiben,

In diesem Sinne konnte im Rahmen dieses Vorhabens aber zumindest prototypisch ein erster eigener Bewertungsansatz für das Tool entwickelt werden. Durch die ergänzende Betrachtung von sogenannten Einheitswurzeltests erscheint dieser Ansatz als relativ robuste Projektionsmethode. Für zukünftige Forschungsaktivitäten zur Entwicklung einer einheitlichen KIS-Fortschreibungsmethode wird daher eine intensiviertere empirische Evaluation und inhaltliche Weiterentwicklung dieses Ansatzes empfohlen,



## 2 Konzeptionelle Grundlagen

### 2.1 Wahl der technischen Basis

Die grundsätzlich in Erwägung gezogenen technischen Alternativen bei der Programmierung des Tools können folgendermaßen kategorisiert werden:

- ▶ MS Office 2013 Anwendungen (in erster Linie Excel unter Verwendung eigener Visual Basic Applikationen),
- ▶ Softwareumgebungen zur numerischen Umsetzung genereller mathematischer Aufgabenstellungen (wie bspw. R, Matlab oder Octave),
- ▶ auf zeitreihenanalytische Untersuchungen spezialisierte ökonomische Softwarepakete (bspw. EViews, Rats oder PcGive).

Die im folgenden Kapitel zusammengefasste methodische Diskussion verdeutlicht, dass im Vergleich zu bisherigen Bewertungsverfahren zukünftig umfassendere Anwendungen induktiv-statistischer Methoden anzustreben sind. Unter Würdigung dieses Aspekts wurden MS Office Anwendungen bereits zu Projektbeginn nicht als dauerhaft zukunftsfähige Umgebung zur Entwicklung der Anwendung angesehen. Während einfache ökonomische Schätzungen auch in dieser Umgebung durchaus implementiert werden können, sind komplexere mathematische Verfahren üblicherweise nicht unmittelbar umsetzbar. Da MS Excel zudem grundsätzlich als ein Tabellenkalkulationsprogramm konzipiert wurde, leiden Analysen komplexerer Datenstrukturen rasch unter einem Verlust an Übersichtlichkeit.

Im Vergleich hierzu sind sowohl generelle mathematische Programmierumgebungen (wie bspw. R) als auch verschiedene auf zeitreihenökonomische Analysen spezialisierte Softwarepakete (wie bspw. EViews) zur Umsetzung der in diesem Gutachten diskutierten Methoden geeignet. Im Rahmen dieses Vorhabens waren zwei Argumente ausschlaggebend für die Entscheidung, die Anwendung in R zu programmieren: Da entsprechende Installationsroutinen für sämtliche gängigen Betriebssysteme (Linux, Mac OS X sowie Windows) kostenfrei und ohne weitere Zugangsbeschränkungen im „Comprehensive R Archive Network“ (CRAN) abgerufen werden können,<sup>12</sup> garantiert die Umsetzung in R die grundsätzlich Lauffähigkeit der Anwendung auf sämtlichen Arbeitsplatz-PCs des Umweltbundesamtes. Zudem verfügte das Umweltbundesamt bereits über eine R-basierte Programmierung dieses Trend-Identifikationsalgorithmus. Mit der Entscheidung für R war somit unmittelbar gewährleistet, dass auch dieses rechenintensive Verfahren, welches auch zum Monitoring der Indikatoren der „Deutsche Anpassungsstrategie an den Klimawandel“ (DAS) genutzt wird (UBA 2015), in der zu entwickelnden Anwendung zur Verfügung stehen würde.

Die konkrete Programmierung erfolgte in der R-Version 3.4.0, wobei RStudio Version 1.0.143 als Entwicklungs- und Testumgebung verwendet wurde.<sup>13</sup>

---

<sup>12</sup> <https://cran.r-project.org/>

<sup>13</sup> Weiterführende Informationen zu RStudio können unter <https://support.rstudio.com/hc/en-us> abgerufen werden.

## 2.2 Wahl der zu implementierenden statistischen Methoden

Eine grundlegende Diskussion der durch das Tool zur Verfügung gestellten Bewertungsverfahren, der dabei durch die Nutzerinnen und Nutzer variierbaren Bewertungsoptionen sowie weitergehender technischer Details (Softwareumgebung, Datenstruktur der Inputdaten, Ausgabeformate, etc.) erfolgte am 26.07.2017 im Rahmen eines halbtägigen Fachgesprächs mit den fachlichen Betreuern des Umweltbundesamtes in Dessau. Zentrale Diskussionspunkte dieses Fachgesprächs können folgendermaßen zusammengefasst werden:

- ▶ Art der Datenerfassung.
- ▶ Technische Basis der Anwendung (Excel, R, ...).
- ▶ Zu berücksichtigende Bewertungsverfahren.
- ▶ Anzuwendende Schätzmethoden und Spezifikationsalgorithmen.
- ▶ Methoden zur Bewertung der Zielerreichung.
- ▶ Auswahlmöglichkeiten bei der Bewertung.
- ▶ Ergebnisausgabe.

Die damit einhergehenden Anwendungsaspekte (wie bspw. implementierte Auswahlmöglichkeiten oder Art und Umfang der Ergebnisausgabe) wurden in einem eigenständigen Benutzerhandbuch dokumentiert. Aus Platzgründen wird daher im Rahmen dieses Abschlussberichts auf eine Detaildokumentation dieser (in erster Linie technischen) Details weitestgehend verzichtet.

Auf Basis der implementierten Bewertungsroutinen zielt das vorliegende Sachverständigengutachten stattdessen auf eine komplementäre Darstellung wesentlicher dabei identifizierter methodischer Aspekte ab. Für die methodische Einordnung der betrachteten statistischen Verfahren wird einer Unterteilung in deskriptive Ansätze, explorative Datenanalysen sowie induktive Analysen gefolgt.<sup>14</sup> Gemäß dieser Unterteilung können, wie nachfolgend argumentiert wird, die derzeit zur Indikatorbewertung angewandten Verfahren als deskriptive Ansätze und explorative Analysemethoden eingeordnet werden.

---

<sup>14</sup> Einleitende Erläuterungen dieser Begriffe sowie allgemeine Anmerkungen zu den unterschiedlichen Ziel-setzungen verschiedener statistischer Methoden können bspw. dem ersten Kapitel von Fahrmeir et al. (2016) entnommen werden.

## 2.2.1 Deskriptive Statistiken

### 2.2.1.1 Zusammenfassende Stichprobeninformationen

Abbildung 5: Beispielhafte Darstellung des Berichtsumfangs der Report-Dateien (Auszug)

```

Report_ABf2_RECL.txt - Editor
Datei Bearbeiten Format Ansicht ?

Die Parameter basieren auf folgender Schätzgleichung:
Time series regression with "numeric" data:
Start = 1, End = 13

Call:
dynlm(formula = trendeq, data = reg_df)

Residuals:
      Min       1Q   Median       3Q      Max
-0.03894 -0.01516 -0.00814  0.02093  0.03483

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.053542   0.013163  307.954 < 2e-16 ***
Zeit         0.011586   0.001862   6.224 0.0000649 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02511 on 11 degrees of freedom
Multiple R-squared:  0.7788,    Adjusted R-squared:  0.7587
F-statistic: 38.74 on 1 and 11 DF, p-value: 0.00006493

Die Projektion folgt der angestrebten Entwicklungstendenz
Gesamtzahl der vorgegebenen Zielwerte: 1
Davon in dieser Stichprobe prüfbar: 1

-----

Berechnung des Zielerreichungsgrades:
Nächstes vorgegebenes Zieljahr: 2020
    
```

Quelle: Eigene Darstellung exemplarischer Befunde bei Anwendung des im Rahmen des Vorhabens entwickelten Tools.

Deskriptive Verfahren dienen zur anwenderfreundlichen Aufbereitung und Komprimierung des Informationsgehalts von Datensätzen. Um die Nutzerinnen und Nutzer innerhalb des UBAs in die Lage zu versetzen, die jeweiligen Befunde bei Anwendungen der Analysealgorithmen im gegebenen Gesamtzusammenhang sinnvoll interpretieren zu können, legt das Tool für jeden Indikator eine umfassende Report-Datei an, welche individuelle deskriptive Statistiken zusammenfasst. Ein beispielhafter Auszug aus einer solchen Report-Datei findet sich in Abbildung 5.

Daneben wird vom Tool beim Programmablauf eine Übersicht sämtlicher analysierter Indikatorreihen im .csv-Format angelegt. Wie der Darstellung in Abbildung 6 entnommen werden kann, enthält diese Darstellung insbesondere eine Übersicht zu den analysierten Zeiträumen sowie den innerhalb dieser Analysezeiträume vorliegenden Beobachtungen. Daneben werden ausgewählte Ergebniszusammenfassungen für die in den folgenden Kapiteln vorgestellten Verfahren ausgewiesen.

Abbildung 6: Beispielhafte Darstellung des Berichtsumfangs der Ergebnisübersicht (Auszug)

File	Index	AngestrebteTender	Datenstand	AnzahlZielwerte	Startjahr	Endjahr	Beobachtungen
Indikatorenset_Paper.xls	KLIM1_THG_GESAMT	Fallend	GWS, Sommer 2017	4	1990	2015	26
Indikatorenset_Paper.xls	ENER04_EES_STROM	Steigend	GWS, Sommer 2017	4	1990	2015	26
Indikatorenset_Paper.xls	MOBI1_EEV_GESAMT	Fallend	GWS, Sommer 2017	3	1991	2014	24
Indikatorenset_Paper.xls	ENER04_EES_EEV	Steigend	GWS, Sommer 2017	3	2000	2015	16
Indikatorenset_Paper.xls	LUFT1_EMIS_MITTEL	Fallend	GWS, Sommer 2017	2	1995	2015	21
Indikatorenset_Paper.xls	TERR3_SUV_MA4TREND	Fallend	GWS, Sommer 2017	2	1996	2015	20
Indikatorenset_Paper.xls	ROHS1_GSPRODKT	Steigend	GWS, Sommer 2017	2	2000	2011	12
Indikatorenset_Paper.xls	ENER01_PREV	Fallend	GWS, Sommer 2017	2	1990	2016	27
Indikatorenset_Paper.xls	ENER02_EEP	Steigend	GWS, Sommer 2017	2	1990	2015	26
Indikatorenset_Paper.xls	ENER05_KWK_STROM	Steigend	GWS, Sommer 2017	2	2003	2015	13
Indikatorenset_Paper.xls	AQUA6_N_NORDSEE_MAS	Fallend	GWS, Sommer 2017	1	1984	2014	31
Indikatorenset_Paper.xls	TERR_SPECIES_GESAMT	Steigend	GWS, Sommer 2017	1	1970	2013	18
Indikatorenset_Paper.xls	GESU_BINNEN_GRENZWS	Steigend	GWS, Sommer 2017	1	1992	2016	25
Indikatorenset_Paper.xls	AQUA6_N_OSTSEE_MAS	Fallend	GWS, Sommer 2017	1	1998	2014	17
Indikatorenset_Paper.xls	AGRI1_STICK	Keine	GWS, Sommer 2017	1	1991	2013	23
Indikatorenset_Paper.xls	WIRT5_EMAS_ORTE	Steigend	GWS, Sommer 2017	1	2005	2016	12
Indikatorenset_Paper.xls	AQUA5_P_GESAMT	Fallend	GWS, Sommer 2017	1	1982	2014	33
Indikatorenset_Paper.xls	GESU_KUESTE_GRENZWS	Steigend	GWS, Sommer 2017	1	1992	2016	25
Indikatorenset_Paper.xls	ROHS2_RMCPC_GESAMT	Fallend	GWS, Sommer 2017	0	2000	2011	12
Indikatorenset_Paper.xls	INDU1_PEV	Keine	GWS, Sommer 2017	0	1995	2014	20
Indikatorenset_Paper.xls	AGRI3_OEKOS	Steigend	GWS, Sommer 2017	0	1996	2015	20
Indikatorenset_Paper.xls	AGRI2_GRLND	Keine	GWS, Sommer 2017	0	1991	2016	26
Indikatorenset_Paper.xls	KONS2_CO2_GESAMT	Keine	GWS, Sommer 2017	0	2000	2013	9
Indikatorenset_Paper.xls	KLIM2_ERWAERMUNG	Keine	GWS, Sommer 2017	0	1850	2016	167
Indikatorenset_Paper.xls	MOBI2_GVK_BINNENSCHIFF	Keine	GWS, Sommer 2017	0	1999	2013	15
Indikatorenset_Paper.xls	KLIM3_HOTDAYS	Keine	GWS, Sommer 2017	0	1951	2016	66

Quelle: Eigene Darstellung exemplarischer Befunde bei Anwendung des im Rahmen des Vorhabens entwickelten Tools.

### 2.2.1.2 Rangkorrelationskoeffizient nach Spearman

Der Rangkorrelationskoeffizient nach Spearman ist eine skalierte Maßzahl zur Beurteilung des monotonen Zusammenhangs zwischen zwei Variablen. Nachdem dieser Koeffizient in der Vergangenheit auch in der Nachhaltigkeitsberichterstattung des Statistischen Bundesamts verwendet wurde und ab 2013 auch Verwendung in der Nachhaltigkeitsberichterstattung von Eurostat fand,<sup>15</sup> wird diese Maßzahl auch im Tool automatisch für die Rangkorrelation zwischen den beobachteten Indikatorwerten und einer im Zeitablauf monoton ansteigenden Trendvariable berechnet.<sup>16</sup> Auf die Berechnung weiterer Konkordanzmaße (Kendalls- $\tau$  oder Goodmans und Kruskals  $\gamma$ -Koeffizient) wurde verzichtet.

Wie Abbildung 7 anhand stilisierter Indikatorzeitreihen verdeutlicht, nimmt dieser Korrelationskoeffizient den Wert Eins an, wenn im Zeitablauf streng monoton ansteigende Indikatorwerte beobachtet werden (linke Abbildungsreihe). Für im Zeitablauf streng monoton fallende Indikatorwerte wird ein Wert von minus Eins ausgewiesen (rechte Abbildungsreihe). Korrelationskoeffizienten nahe Null weisen darauf hin, dass der Indikator im Zeitablauf keine einheitliche Monotonieeigenschaft aufweist (untere Abbildungen der mittleren Spalte). Für im Zeitablauf konstante Indikatorwerte (obere Abbildung der mittleren Reihe) sowie bei Vorliegen von weniger als fünf Beobachtungen wird diese Maßzahl nicht berechnet.

<sup>15</sup> Siehe hierzu Eurostat (2014).

<sup>16</sup> Siehe bspw. auch BMUB (2015) für weitere Anwendungen des Rangkorrelationskoeffizienten nach Spearman.

Abbildung 7: Rangkorrelationskoeffizienten für alternative Trendverläufe

**Streng monoton ansteigende Beobachtungen**  
Rangkorrelationskoeffizient=1



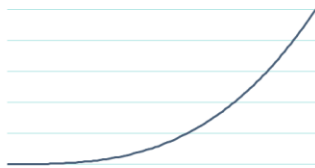
**Invariante Indikatorbeobachtungen**  
Rangkorrelationskoeffizient nicht anwendbar



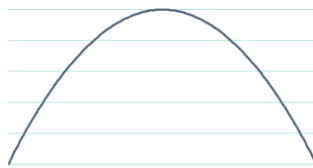
**Streng monoton abnehmende Beobachtungen**  
Rangkorrelationskoeffizient=-1



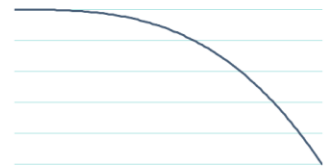
**Streng monoton ansteigende Beobachtungen**  
Rangkorrelationskoeffizient=1



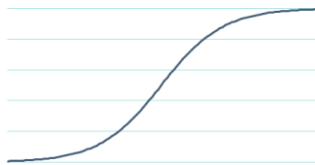
**Variierende Monotonieeigenschaften**  
Rangkorrelationskoeffizient (im Extremfall)=0



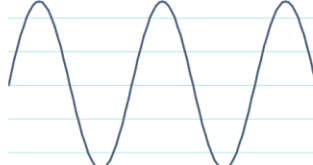
**Streng monoton abnehmende Beobachtungen**  
Rangkorrelationskoeffizient=-1



**Streng monoton ansteigende Beobachtungen**  
Rangkorrelationskoeffizient=1



**Variierende Monotonieeigenschaften**  
Rangkorrelationskoeffizient (im Extremfall)=0



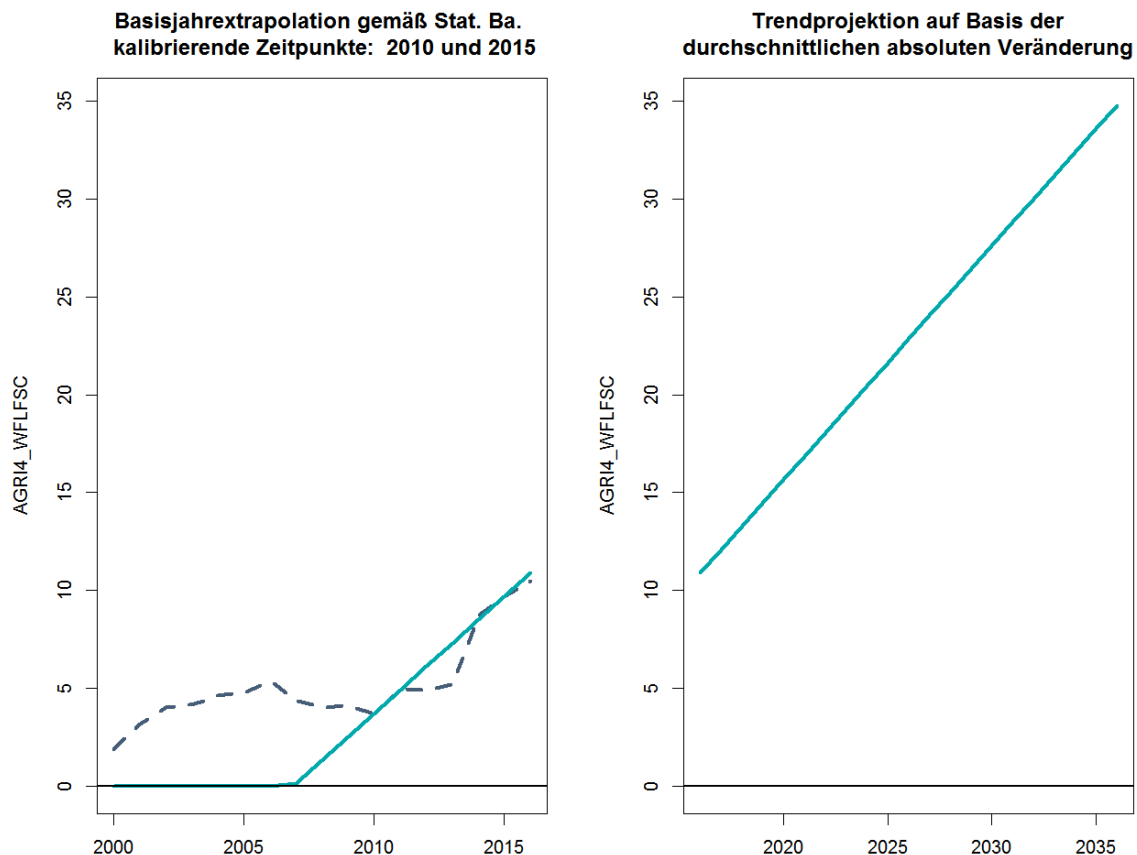
**Streng monoton abnehmende Beobachtungen**  
Rangkorrelationskoeffizient=-1



Quelle: Eigene schematische Darstellung

### 2.2.1.3 Trendidentifikation und -extrapolation gemäß Methode des Statistischen Bundesamtes

**Abbildung 8: Beispiel einer Trendextrapolation gemäß Methode des Statistischen Bundesamtes**



Quelle: Beispielhafte eigene Berechnung der mittleren absoluten Veränderung zwischen den Jahren 2015 und 2010 (linke Abbildung) sowie eigene Projektion der resultierenden Trendgeraden bis zum Jahr 2035 (rechte Abbildung).

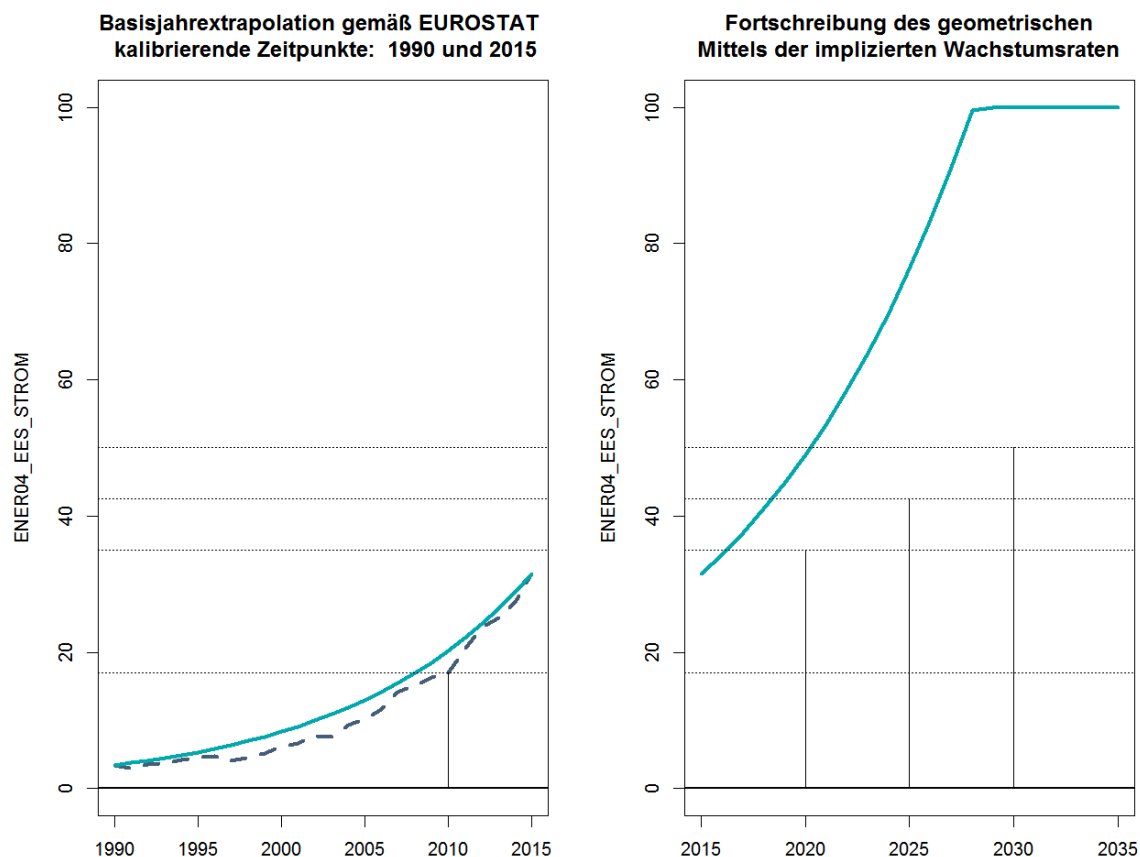
Zur Extrapolation zukünftiger Entwicklungen werden in der Nachhaltigkeitsberichterstattung des Statistischen Bundesamtes die sechs jüngst verfügbaren Beobachtungen eines Indikators verwendet (siehe hierzu bspw. auch Blumers und Kaumanns 2017, S. 102). Auf Basis dieser Beobachtungen wird die durchschnittliche jährliche Veränderung des analysierten Indikators innerhalb der vergangenen fünf Jahre berechnet. In der Extrapolation wird daraufhin die weitere Entwicklung des betrachteten Indikators unter Beibehaltung dieser durchschnittlichen jährlichen Veränderung abgebildet. Abbildung 8 verdeutlicht, dass auf dieser Basis auch eine Trendfortschreibung als Extrapolation einer Geraden durch die historischen Beobachtungen der Indikatorwerte im gewählten Basisjahr (2015) sowie dem frühesten bei dieser Bewertung berücksichtigten Beobachtungsjahr generiert werden kann.

Entsprechende Berechnungen werden im Tool automatisch für jeden Indikator durchgeführt, wobei die Anwenderinnen und Anwender das Basisjahr dieser Berechnungen (also das Jahr, welches als letzte historisch verfügbare Beobachtung verwendet wird) flexibel auswählen können. Sofern der Indikator in diesem Zeitraum fehlende Werte aufweist, werden vom Tool automatisch frühere Beobachtungen zur Berechnung der durchschnittlichen jährlichen

Veränderung auf Basis der jüngsten sechs verfügbaren Beobachtungen bis zum gewählten Basisjahr herangezogen. Sollte sich dabei herausstellen, dass für das gewählte Basisjahr weniger als vier verwendbare Zeitpunkte zur Verfügung stehen, wird von einer entsprechenden Berechnung abgesehen.

#### 2.2.1.4 Trendidentifikation und -extrapolation gemäß Eurostat

Abbildung 9: Beispiel einer Trendextrapolation gemäß Eurostat-Methode



Quelle: Beispielhafte eigene Berechnung des geometrischen Mittels des beobachteten Wachstums zwischen den Jahren 1990 und 2015 (linke Abbildung) sowie Projektion des resultierenden Wachstumspfad bis zum Jahr 2035.

Eurostat wendet in der Indikatorberichterstattung ein Verfahren an, bei welchem in der Projektionsperiode von einer Beibehaltung des in der Historie für einen ausgewählten Indikator zwischen zwei Zeitpunkten durchschnittlich beobachteten jährlichen Wachstums ausgegangen wird. Das durchschnittliche Wachstum wird dabei als geometrisches Mittel aus dem Wachstumsfaktor der beiden gewählten Beobachtungen berechnet (Eurostat 2014).

Während Eurostat diese Methode stets für fest vorgegebene Zeitintervalle anwendet (5 oder 10 Jahre), erlaubt die im Tool erfolgte Umsetzung eine Anwendung für flexibel definierte Zeiträume. So zeigt bspw. Abbildung 9 zeigt die Befunde bei Anwendung dieser Methode auf den beispielhaft ausgewählten Indikator „Anteil erneuerbarer Energien am Bruttostromverbrauch“ im Zeitraum 1990 bis 2015. Im linken Teil der Abbildung wird der zur Berechnung verwendete „Schätz“-Zeitraum dargestellt. Wie leicht zu erkennen ist, wurden in diesem Beispiel sämtliche

verfügbaren Beobachtungen zur Berechnung der implizierten Wachstumsraten verwendet (für die erste sowie die letzte historische Beobachtung stimmen historische Beobachtung sowie der stilisierte Trendverlauf exakt überein).

## 2.2.2 Explorative Datenanalysen: Trendschätzungsverfahren „DAS Monitoring“

### 2.2.2.1 Vorstellung des implementierten Algorithmus

Wie bereits einleitend festgehalten wurde, zeichnet sich das Trendschätzungsverfahren „DAS Monitoring“ im Vergleich zu den bislang betrachteten Indikatoranalyse-Ansätzen durch einen umfassenden Einsatz zeitreihenanalytischer Methoden aus. Da dem UBA bereits aus einem früheren Vorhaben zur indikatorbasierten Berichterstattung (Schönthaler und von Andrian-Werburg 2015) lauffähige R-Skripte zur Anwendung dieses Verfahrens vorlagen, wurden diese Routinen vollständig in die im Rahmen dieses Vorhabens entwickelte R-Anwendung integriert.<sup>17</sup> Das Tool ist damit in der Lage, für jeden vorliegenden Indikator die historische Anpassungsgüte der in Abbildung 10 stilisiert dargestellten Trendpolynome zu testen.

Die Grundidee dieses Analyseansatzes, welcher insbesondere auf eine explizite Schätzung der in zeitreihenbasierten Regressionsanalysen üblicherweise anzutreffenden Autokorrelationsstrukturen abzielt, wurde im Kontext der Indikatorberichterstattung von Martone (2013) für den Arbeitskreis Umweltökonomische Gesamtrechnungen der Länder (UGRdL) entwickelt. Eine vollständige Übersicht der dabei angewandten Modell-Selektionsschritte kann Abbildung 11 entnommen werden. Wie aus dieser Abbildung ersichtlich wird, kennzeichnet insbesondere die explizite mathematische Modellierung von (signifikant geschätzten) Autokorrelations-Symptomen diesen Algorithmus. Dieses Merkmal reflektiert den Befund, dass bei Anwendung zeitreihenanalytischer Verfahren häufig autokorrelierte Residuen identifiziert werden. Um numerische Verzerrungen der geschätzten Parameter zu vermeiden, wird bei Vorliegen signifikanter Autokorrelationseffekte, in der Tradition von Cochrane und Orcutt (1949), eine explizite Schätzung der Autokorrelationsstruktur der Residuen durchgeführt. Zur numerischen Identifikation der Parameter werden iterative Maximum-Likelihood-Schätzer verwendet.

Die im Rahmen dieses Vorhabens genutzten R-Skripte wurden zur Verstetigung der indikatorbasierten DAS-Berichterstattung am Statistischen Beratungslabor der Ludwig-Maximilians-Universität (LMU) entwickelt. Diese Monitoring Routinen orientieren sich umfassend an den in Abbildung 11 dargestellten Prozessabläufen. Allerdings wird dabei eine Auswahl höhergradiger Trendpolynome tendenziell befördert, indem „... bei Datenreihen mit mehr als 12 Datenpunkten grundsätzlich zunächst quadratische Trends betrachtet werden.“ (Schönthaler und von Andrian-Werburg 2015, S. 53).

Da zur Abbildung autokorrelierter Residuen maximal AR(2)-Prozesse geschätzt werden, lässt sich diese zeitreihenanalytische Modellierung in ihrer generellsten Form folgendermaßen darstellen:

$$(1.a) \quad y_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + u_t,$$

$$(1.b) \quad u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t,$$

---

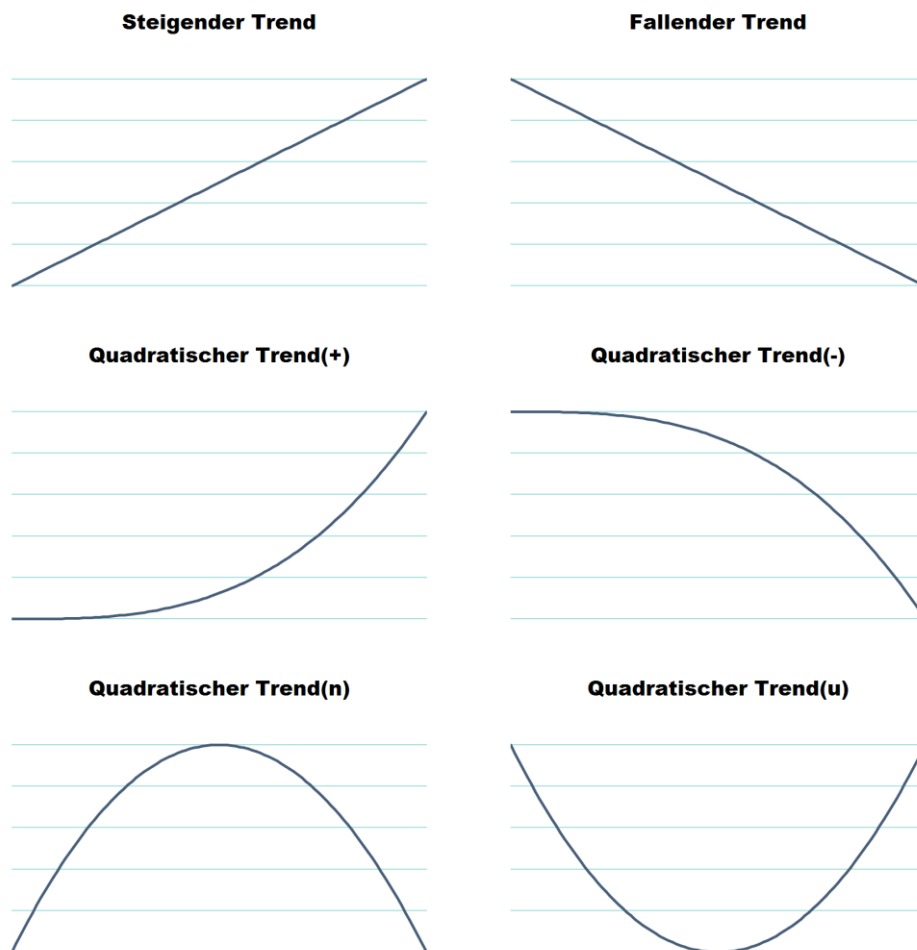
<sup>17</sup> Autor dieses Programmcodes (datierend vom Sommer 2014) ist Shuai Shao.



$$(1.c) \quad \epsilon_t \sim \text{iid } N(0, \sigma^2).$$

Dabei bezeichnet  $y_t$  die beobachtete Ausprägung des analysierten Indikators zum Zeitpunkt  $t$ . Die Abweichung dieser Beobachtung vom unterstellten Trendterm  $\gamma_0 + \gamma_1 t + \gamma_2 t^2$  ist mit  $u_t$  bezeichnet. Die Realisation eines unabhängig identisch normalverteilter Zufallsterms zum Zeitpunkt  $t$  wird mit  $\epsilon_t$  bezeichnet.

**Abbildung 10: Durch das Trendschätzungsverfahrens „DAS Monitoring“ geprüfte Trendmuster**

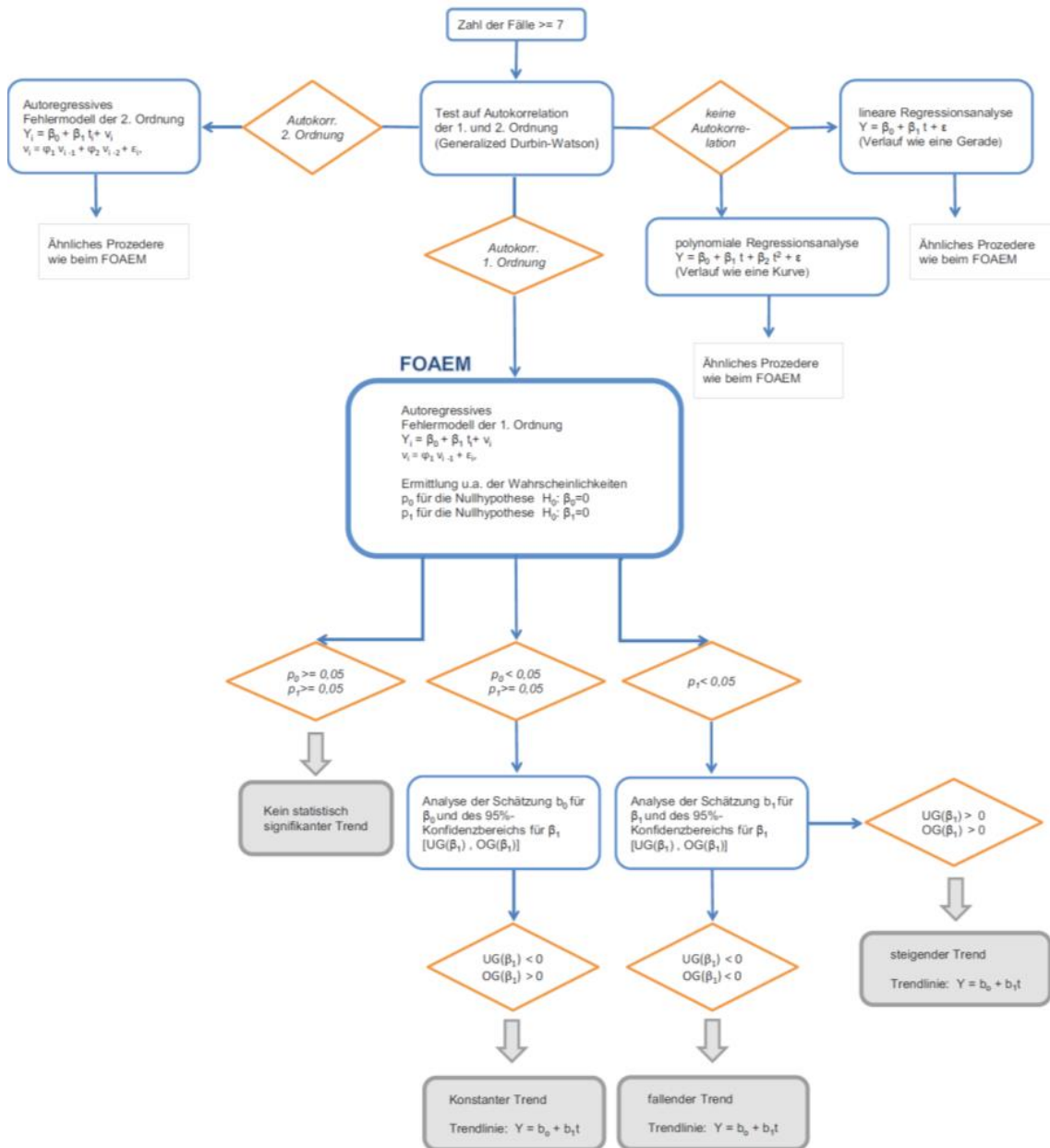


Quelle: Eigene stilisierte Darstellung.

Da die grundsätzlichen Vorgehensweisen zur Schätzung der jeweiligen Trendverläufe sowohl von Martone (2013) als auch in Schönthaler et al. (2015) hinreichend dokumentiert wurden (siehe hierzu beispielhaft Abbildung 11), wird auf eine weitergehende Reproduktion entsprechender Beschreibungen an dieser Stelle verzichtet. Es sei allerdings darauf hingewiesen, dass (ähnlich wie bei den vorherigen Trend-Fortschreibungsmethoden in Anlehnung an Vorgehensweisen des Statistischen Bundesamts sowie von Eurostat) auch bei der Integration dieser Routinen eine möglichst flexible Anwendbarkeit angestrebt wurde. So ist es

den Nutzerinnen und Nutzern bspw. möglich, die zu analysierenden Zeiträume individuell für jeden Indikator einzeln vorzugeben.<sup>18</sup>

Abbildung 11: Grundstrukturen des Schätzalgorithmus von Martone (2013)



Quelle: Martone (2013, S. 86).

### 2.2.2.2 Ergänzende Anmerkungen

Das soeben vorgestellte Trendbewertungsverfahren ermöglicht eine systematische Beschreibung historischer Zeitreihenbefunde. Da der umfangreiche Identifikations- und

<sup>18</sup> Entsprechende Details können dem zugehörigen Benutzerhandbuch entnommen werden.

Schätzalgorithmus eine automatisierte Klassifikation vergangener Indikatorbeobachtungen ermöglicht, ist der Leistungsumfang dieses Verfahrens im Vergleich zu den zuvor betrachteten deskriptiven Ansätzen als wesentlich weitreichender zu beurteilen. Da dieses Trendschätzungsverfahren zudem im Rahmen des DAS Monitorings regelmäßig zur Bewertung historischer Indikatorentwicklungen angewandt wird, wurden die verfügbaren R-Skripte zur Anwendung dieses Verfahrens auf Wunsch des Auftraggebers im Tool implementiert.

Für die im Rahmen dieses Vorhabens zu erstellende R-Anwendung wurde in der Leistungsbeschreibung allerdings explizit gefordert, dass diese die analysierten Indikatoren nicht nur hinsichtlich ihres bisherigen Verlaufs beschreibt, sondern auch die *Signifikanz* identifizierter Trendverläufe abschätzt sowie auf Basis dieser Abschätzungen eine *Projektion zukünftiger Zielerreichungsgrade* ermöglicht. Da die DAS Routinen allerdings nicht für Projektionszwecke entwickelt wurden, ist die Integration dieser Routinen im Tool als eine explorative Vorgehensweise aufzufassen. Angesichts der in Kapitel 3 diskutierten formalstatistischen Aspekte erscheinen diesbezüglich weitergehende methodische Forschungsarbeiten notwendig. Nachdem entsprechende Forschungsaktivitäten im Rahmen dieses Vorhabens nicht intendiert waren, wird eine vertiefte empirische Evaluation dieser Aspekte in zukünftigen eigenständigen Forschungsprojekten empfohlen.

### 2.3 Beurteilung zukünftig zu erwartender Zielerreichungsgrade

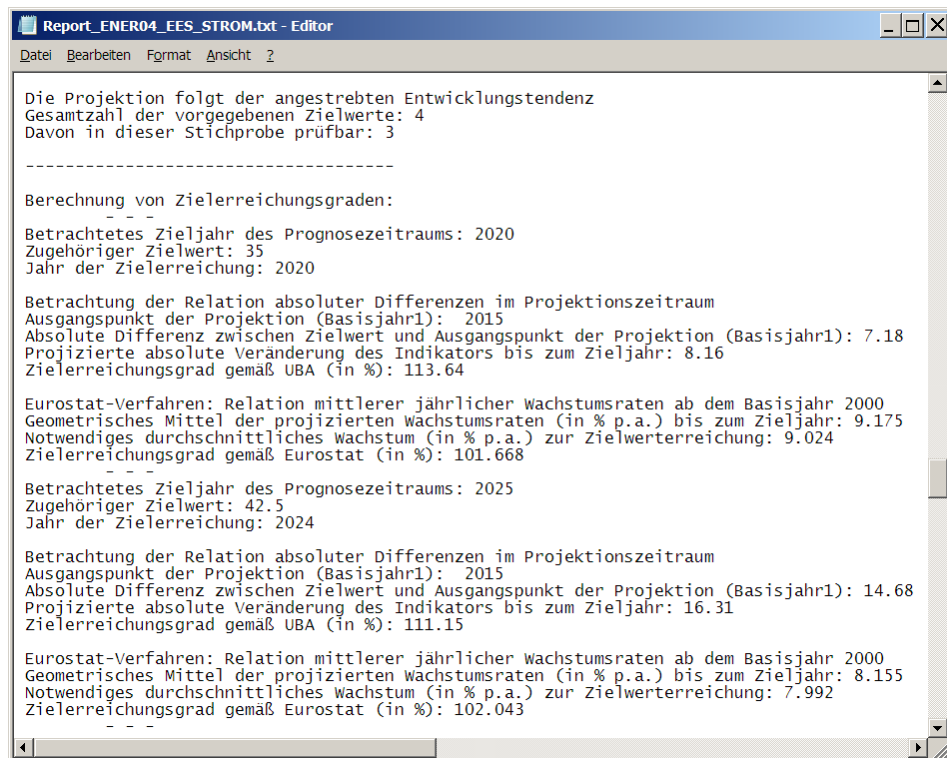
Die im Rahmen dieses Vorhabens zu erstellende R-Anwendung sollte die analysierten Indikatoren nicht nur hinsichtlich ihres bisherigen Verlaufs beschreiben, sondern auch eine Projektion zukünftiger Zielerreichungsgrade ermöglichen. Die diesbezüglich in den jeweiligen Reportdateien zur Verfügung gestellten Informationen veranschaulicht Abbildung 12.<sup>19</sup>

Wenn für einen Indikator eine langfristig angestrebte Entwicklungstendenz („Fallend“, bzw. „Steigend“) vorgegeben wurde, erfolgt zu Beginn dieses Ergebnisreports zunächst eine qualitative Beurteilung der langfristig projizierten Entwicklungstendenz (obere Zeile des Screenshots). Nach einer kurzen Information zur Anzahl der insgesamt verfügbaren Zielwerte sowie der auf Basis der vorliegenden Projektionen prüfbareren Zielwerte erfolgt dann für sämtliche im Projektionszeitraum überprüfbarere Zieljahre (im hier gezeigten Beispiel 2020 und 2025) eine detaillierte Beurteilung der Zielerreichungsgrade. Nachrichtlich wird dabei zunächst darüber berichtet, ob (bzw. in welchem Jahr) der betrachtete Zielwert in der analysierten Projektion tatsächlich erreicht wird. Danach erfolgt eine Zusammenfassung detaillierter Berechnungsergebnisse für zwei unterschiedlich definierte Zielerreichungsgrade.

---

<sup>19</sup> Für ergänzende Hinweise auf entsprechende Zusammenfassungen in der übergeordneten „Results“-Datei wird wiederum auf das Benutzerhandbuch der Anwendung verwiesen.

Abbildung 12: Beispiel projizierter Zielerreichungsgrade



Quelle: Eigene Darstellung auf Basis einer exemplarischen Anwendung der R-Anwendung.

Der im Rahmen dieses Vorhabens entwickelte eigene „Zielerreichungsgrad gemäß UBA“ setzt dabei die projizierte absolute Veränderung des betrachteten Indikators zwischen einem frei wählbaren Basisjahr („Zielerreichungsgrad Basisjahr“, in diesem Beispiel das Jahr 2015) sowie dem vorgegebenen Zieljahr in Relation zu der für eine Einhaltung des Zielwertes notwendigen absoluten Veränderung dieses Indikators im selben Zeitraum. Im Vergleich hierzu betrachtet der „Zielerreichungsgrad gemäß Eurostat“ die durchschnittliche jährliche Wachstumsrate zwischen einem (wiederum individuell wählbaren) Basisjahr („Basisjahr2“, in diesem Beispiel das Jahr 2000) sowie dem vorgegebenen Zieljahr in Relation zu der für eine Einhaltung des Zielwertes in diesem Zeitraum durchschnittlich notwendigen jährlichen Wachstumsrate.<sup>20</sup>

Auf eine weitergehende qualitative Klassifikation dieser Befunde (bspw. in Form der in Abbildung 12 dargestellten Wettersymbole) wird im Tool bewusst verzichtet. Diese Entscheidung ist dadurch begründet, dass die Festlegung entsprechender Klassifikationsstufen sowie der zur Einordnung individueller Projektionsergebnisse zu den jeweiligen Bewertungsklassen anzuwendenden Schwellenwerte nicht objektiv begründbar ist. Um entsprechende Einordnungen zukünftig flexibel vornehmen zu können, wurde daher in Abstimmung mit den fachlichen Betreuern des Umweltbundesamtes beschlossen, dass das Tool sämtliche hierfür notwendigen Berechnungsergebnisse zur Verfügung stellen soll. Die Etablierung eines eigenen Klassifikationsschemas sowie dessen Anwendung auf Basis der durch das Tool verfügbaren Informationen verbleibt daher für zukünftige Forschungsaktivitäten.

<sup>20</sup> Ergänzend sei angefügt, dass abschließend auch eine zusammenfassende Übersicht sämtlicher Zieljahre und -werte sowie deren Gegenüberstellung mit jeweils projizierten Indikatorwerten erfolgt. Im Ergebnisauszug der obigen Abbildung kann diese Funktionalität nicht erkannt werden, weitere Details hierzu werden aber im Tool-Handbuch kurz angesprochen.

**Abbildung 13: Bewertung von Zielerreichungsgraden, Beispiel der Nachhaltigkeitsberichterstattung des Statistischen Bundesamtes.**

---



Bei Fortsetzung der Entwicklung beträgt die Abweichung vom Zielwert weniger als 5 % der Differenz zwischen Zielwert und aktuellem Wert;



Bei Fortsetzung der Entwicklung würde das Ziel voraussichtlich um mindestens 5 %, aber maximal um 20 % der Differenz zwischen Zielwert und aktuellem Wert verfehlt;



Der Indikator entwickelt sich zwar in die gewünschte Richtung auf das Ziel zu, bei Fortsetzung der Entwicklung würde das Ziel im Zieljahr aber um mehr als 20 % verfehlt;



Der Indikator entwickelt sich nicht in die gewünschte Richtung und somit vergrößert sich der Abstand zum Ziel.

Quelle: Statistisches Bundesamt (2017, S. 118).

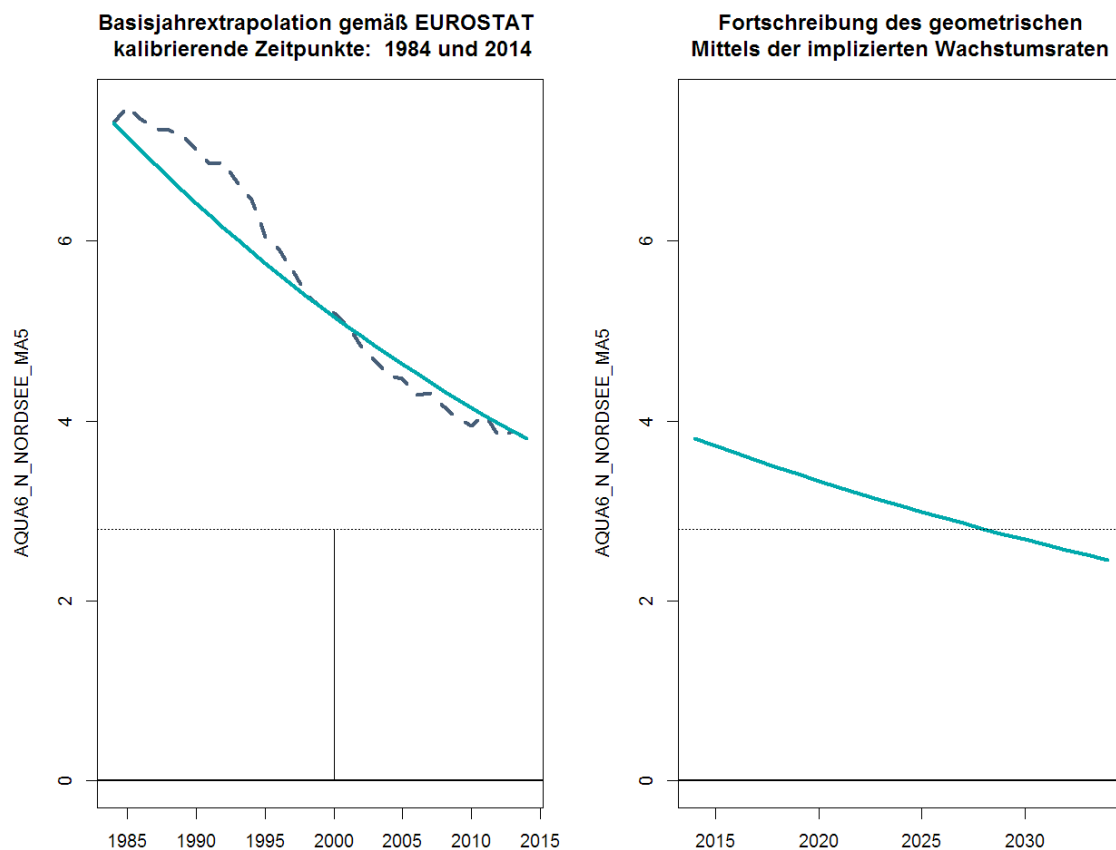
### 3 Robuste Projektionen zukünftiger Trendverläufe: Identifizierte methodische Ansatzpunkte

Dieses Kapitel dient der Vorstellung ausgewählter Ansatzpunkte, welche im Zuge zukünftiger methodischer Weiterentwicklungen vertieft diskutiert werden sollten. Im Anschluss an die in Unterkapitel 3.1 erfolgende zusammenfassende Darstellung der aus Sicht des Autors vordringlich anzusprechenden methodischen Aspekte werden in Abschnitt 3.1.1 naheliegende Optionen zur formal-statistischen Erweiterung der Trendextrapolationsmethode von Eurostat vorgestellt. Abschnitt 3.1.2 verweist auf methodische Aspekte, welche bei üblichen Anwendungen des DAS Monitoring Verfahrens zunächst vernachlässigbar erscheinen, denen im Kontext der Trendextrapolation allerdings eine umfassende Relevanz zugesprochen werden kann. Nachdem im Tool bereits für einige der hier angesprochenen Aspekte zumindest prototypische Bewertungsansätze umgesetzt werden konnten, erfolgt anschließend in Unterkapitel 3.2 eine zusammenfassende Darstellung dieser eigenständigen Entwicklungen.

#### 3.1 Theoretische Anmerkungen

##### 3.1.1 Trendextrapolationen in Anlehnung an Eurostat

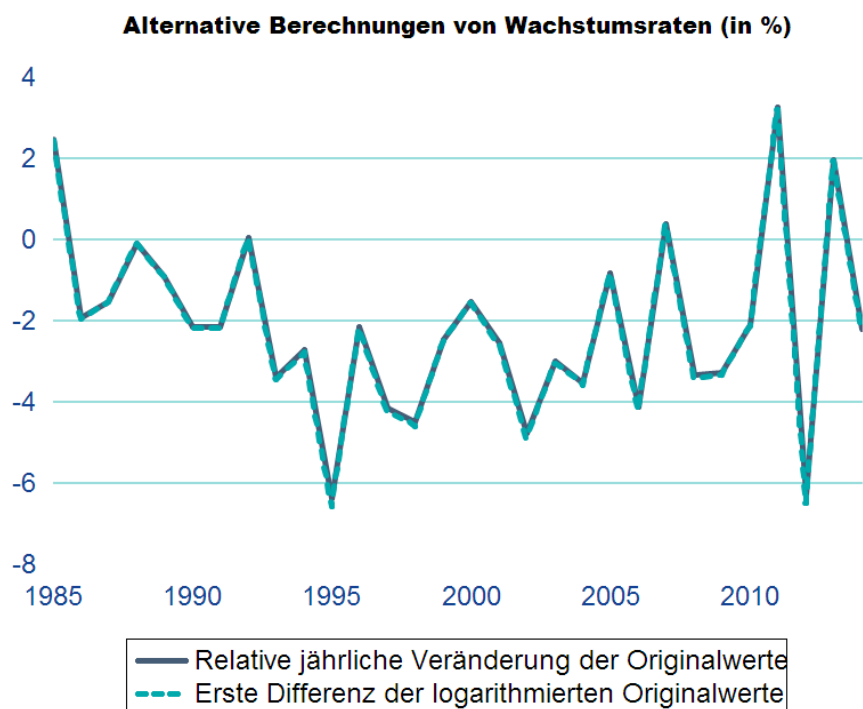
Abbildung 14: Beispiel einer Trendextrapolation in Anlehnung an die Eurostat-Methode



Quelle: Eigene Berechnung des Trendverlaufs für den Indikator „Stickstoffkonzentration der Hauptflüsse zur Nordsee“ zwischen 1984 und 2014 (linke Abbildung) sowie Projektion des resultierenden Indikatorverlaufs bis 2034.

Abbildung 14 betrachtet das Beispiel einer Trendfortschreibung in Anlehnung an die Extrapolationsmethode von Eurostat. Dargestellt sind die historischen Beobachtungen des Indikators „Konzentration Gesamtstickstoff der Hauptflüsse zur Nordsee in Milligramm/Liter“ (gestrichelte Linie im linken Diagramm) sowie der resultierende Trendverlauf bei Berechnung des geometrischen Mittels des zwischen den Jahren 1984 und 2014 berechneten Wachstumsfaktors dieses Indikators (durchgezogene Linie). Der so berechnete durchschnittliche Wachstumsfaktor entspricht einem Wert von ca. 0,978. Die Gesamtstickstoff-Konzentration der Hauptflüsse zur Nordsee wurde also zwischen 1984 und 2014 Durchschnitt um ca. 2,2% pro Jahr reduziert.

**Abbildung 15: Beispiel alternativer numerischer Berechnungen von Wachstumsraten**



Quelle: Eigene Berechnungen auf Basis der Indikatorbeobachtungen der Abbildung 14.

Während der vom Tool in dieser Anwendung projizierte Trendverlauf nicht unwahrscheinlich erscheint, stellt eine übliche Anwendung dieser Methode keine weiteren statistischen Informationen zur Verfügung, um die Plausibilität der hier unterstellten Hypothese („die jährliche Veränderung des Indikators wird zukünftig dem Durchschnitt der in der Vergangenheit beobachteten jährlichen Veränderungen entsprechen) formal zu überprüfen.

Entsprechende Teststatistiken lassen sich allerdings leicht auf Basis einer zusätzlichen Regressionsanalyse berechnen: Bezeichne  $y_t$  die logarithmierten Beobachtungen eines beliebigen Indikators, so entsprechen die ersten Differenzen dieser logarithmierten Beobachtungen

$$\Delta y_t = y_t - y_{t-1}$$

näherungsweise den jeweiligen Wachstumsraten der Indikatorbeobachtungen zwischen den beiden betrachteten Zeitpunkten  $t$  und  $t - 1$ .



Für das soeben betrachtete Beispiel wird dieser Zusammenhang durch Abbildung 15 verdeutlicht. Die durchgezogene graue Linie repräsentiert dabei die (mit dem Faktor 100 multiplizierten) jährlichen Wachstumsraten des Indikators „Konzentration Gesamtstickstoff der Hauptflüsse zur Nordsee“, berechnet als relative Veränderung der Indikatorbeobachtungen für zwei aufeinanderfolgende Jahre. Die gestrichelte grüne Linie repräsentiert die (wiederum mit dem Faktor 100 multiplizierten) jährlichen Wachstumsraten dieses Indikators, berechnet als erste Differenz der logarithmierten Indikatorbeobachtungen. Die numerischen Abweichungen zwischen den jeweils berechneten Werten sind augenscheinlich vernachlässigbar, da beide Kurvenverläufe nahezu identisch erscheinen.

Folglich kann die bei Durchführung einer Trendextrapolation in Anlehnung an das Eurostat-Verfahren unterstellte Hypothese einer im Analysezeitraum konstanten durchschnittlichen Wachstumsrate auch anhand einer Regression der ersten Differenzen der logarithmierten Indikatorbeobachtungen auf eine Konstante überprüft werden. Die entsprechende Testgleichung lässt sich folgendermaßen notieren

$$(2) \Delta y_t = \beta_0 + \varepsilon_t$$

wobei  $\varepsilon_t$  wiederum die Ausprägungen eines statistischen Zufallsterms zum Zeitpunkt  $t$  bezeichnet (vgl. hierzu auch Gleichung 1.c). Die zu schätzende konstante Wachstumsrate ist mit  $\beta_0$  bezeichnet. Gleichung (2) kann mit der üblichen Methode der Kleinsten-Quadrate geschätzt werden, woraufhin die üblichen formalen Fehlspezifikationstests angewandt werden können.

Im hier betrachteten Beispiel resultiert die Schätzung von Gleichung (2) in einem (gerundeten) Schätzwert  $\hat{\beta}_0$  in Höhe von -0.022. Die Schätzung bestätigt also den bereits zuvor berichteten Befund, dass die Gesamtstickstoff-Konzentration der Hauptflüsse zur Nordsee zwischen 1984 und 2014 durchschnittlich um ca. 2,2% pro Jahr reduziert wurden. Darüber hinaus weisen die Regressionsergebnisse aber zudem darauf hin, dass dieser Trendeffekt als deutlich signifikant eingeordnet werden kann (t-Statistik: -5.17) und allgemeine Fehlspezifikationstest keine offensichtlichen Anzeichen aufzeigen, die gegen eine Verwendung dieser Schätzgleichung für Projektionszwecke sprechen würden.

Um entsprechende Analysen auch im Tool anwenden zu können, wurde ein eigener prototypischer Bewertungsalgorithmus entwickelt, welcher eine automatisierte Berechnung der natürlichen Logarithmen sämtlicher eingelesener Indikatorbeobachtungen ermöglicht, um Trendanalysen auf Basis der ersten Differenzen dieser Zeitreihen vorzunehmen. Weiterführende Anmerkungen hierzu finden sich in Kapitel 3.2.

### 3.1.2 Trendextrapolationen in Anlehnung an DAS Monitoring Verfahren

#### 3.1.2.1 Herausforderungen bei der Fortschreibung zeitvarianter Trendeffekte

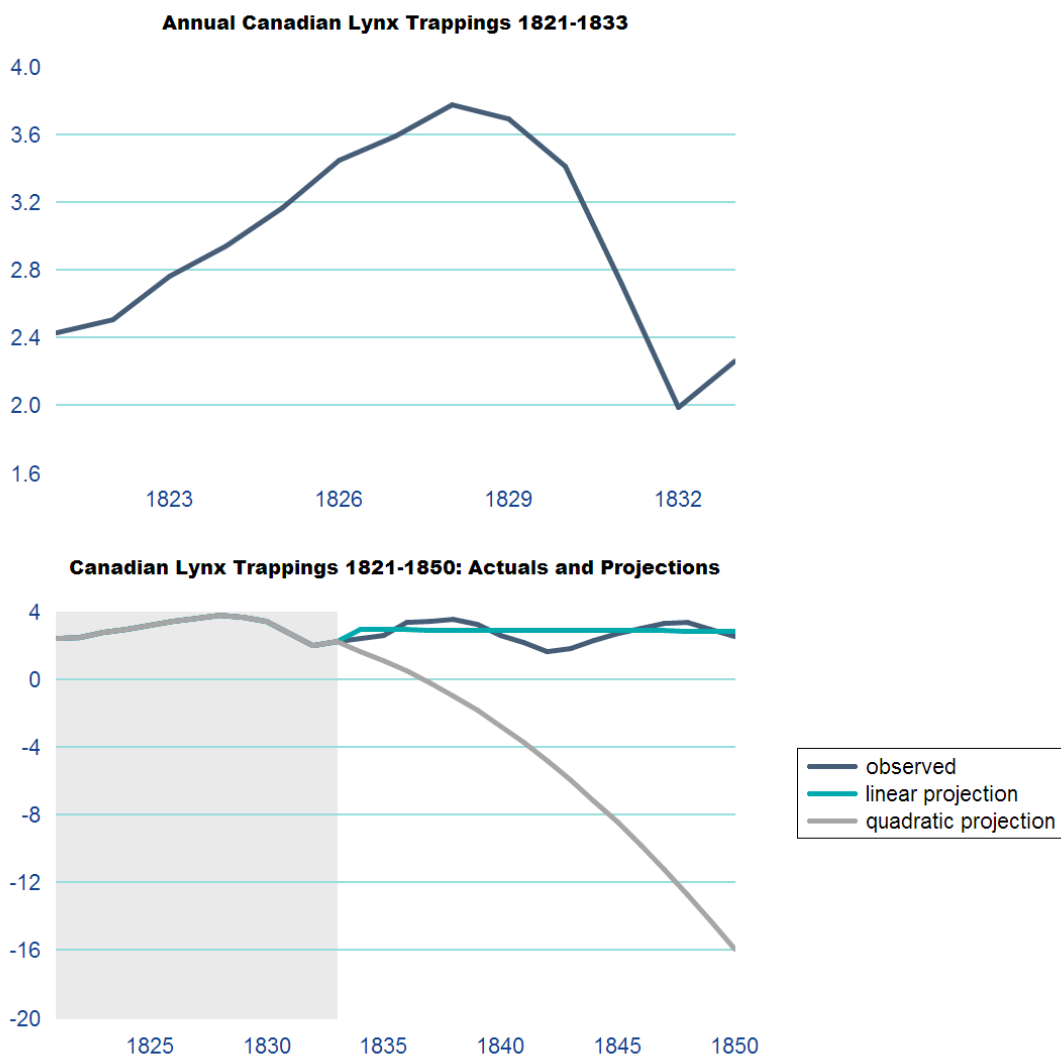
Es ist allgemein bekannt, dass sich jeder gegebene Zeitreihenverlauf durch Polynome höherer Ordnung mit beliebiger Genauigkeit approximieren lässt. Zudem ist aber auch bekannt, dass höhergradige Polynomschätzungen in der Regel keine guten Prognoseeigenschaften aufweisen: „Polynome höherer Ordnung sind außerhalb des Datenbereichs sehr instabil [...] und sind deshalb für Prognosezwecke ungeeignet.“ (Fahrmeir et al. 2016, S. 512).

Abbildung 16 veranschaulicht diesen Zusammenhang am Beispiel einer eigenen empirischen Anwendung. Hierzu wird die in zeitreihenanalytischen Lehrbüchern aufgrund ihrer



dynamischen Eigenschaften häufig als Referenz herangezogene Zeitreihe der jährlich im Gebiet des MacKenzie Rivers (Kanada) gefangenen Luchse betrachtet. Für die obere Grafik wurden aus dieser Zeitreihe die logarithmierten historischen Beobachtungen der Jahre 1821 bis 1833 zusammengestellt. Für diese 13 Beobachtungen ist zu erkennen, dass die ersten acht Beobachtungen durch einen jährlichen Anstieg der Luchszahlen gekennzeichnet sind, woraufhin im Zeitablauf vier Beobachtungen mit jeweils absteigenden Fangzahlen folgen. Zum Ende des Beobachtungszeitraums steigen die Fangzahlen wieder leicht an, erreichen damit aber noch nicht das Ausgangsniveau des Jahres 1821.

**Abbildung 16: Beispielhafte Extrapolation eines geschätzten Trendpolynoms**



Quelle: Eigene Darstellung der Ergebnisse einer Anwendung des R-Skripts zu Schönthaler et al. (2015).

Der Rückgriff auf das hier gewählte Lehrbuchbeispiel ermöglicht es nun, die tatsächliche Performance unterschiedlicher Bewertungs- und Projektionsansätze im Vergleich zur tatsächlich später beobachtbaren historischen Entwicklung zu betrachten. In der unteren Grafik der Abbildung wurde hierzu der Beobachtungszeitraum auf die Jahre 1821 bis 1850

ausgeweitet. Die in dieser Darstellung grau schattierte Fläche kennzeichnet dabei den ursprünglichen Beobachtungszeitraum des oberen Zeitreihenplots (1821-1833). Der historische Verlauf der Zeitreihe wurde wiederum dunkelgrau abgebildet. Die langfristig beobachtbaren Zeitreihenrealisationen sind demnach durch relativ gleichmäßige Schwankungen um einen langfristigen nahezu stabilen Mittelwert charakterisiert.

Im Gegensatz hierzu kennzeichnet ein deutlicher rückläufiger Trend die hellgraue Kurve. Diese Kurve repräsentiert Fortschreibungsergebnisse eines über den Zeitraum 1821-1833 unter Anwendung des DAS Monitoring-Skripts geschätzten Trendpolynoms. Die Fortschreibung erfolgte statisch (also ohne Berücksichtigung zukünftiger Auswirkungen des Fehlerterms) auf Basis der historischen Beobachtung des Jahres 1833.

Zu Vergleichszwecken wurde für die Beobachtungen der Jahre 1821-1833 ebenfalls ein einfacher linearer Trend unter Anwendung der Methode der Kleinsten Quadrate geschätzt. Bezeichne  $y_t$  die logarithmierten Luchsfangzahlen eines beliebigen Jahres  $t$ , so lässt sich die Hypothese, dass diese im Beobachtungszeitraum lediglich zufällig von einem linearen Trend abweichen folgendermaßen notieren:<sup>21</sup>

$$(3) y_t = \gamma_0 + \gamma_1 t + \epsilon_t.$$

Auf Basis dieser Schätzung wäre der weitere Verlauf der (logarithmierten) Luchs-Zeitreihe anhand des grünen Kurvenverlaufs projiziert worden. Der Vergleich dieser Kurve mit der zuvor vorgestellten Fortschreibung der DAS Monitoring-Ergebnisse veranschaulicht, dass aus der historischen Anpassungsgüte eines höhergradigen Trendpolynoms für gegebene Zeitreihenbeobachtungen (welche im hier betrachteten Beispiel durch die Bewertungsroutinen des DAS Monitorings bestätigt wird) keinesfalls auch eine Eignung dieses Trendpolynoms zur zukünftigen Fortschreibung der Zeitreihenentwicklung geschlussfolgert werden kann. Die historischen Schwankungen der analysierten Zeitreihe können durch die Schätzergebnisse der Gleichung (3) nicht reproduziert werden. Allerdings ist offensichtlich, dass diese Gleichung eine wesentlich robustere und damit auch für trendmäßige Fortschreibungen (zumindest besser) geeignete Abschätzung der allgemeinen Lageparameter der Zeitreihe bietet.

Unter Verweis auf dieses Beispiel muss die Eignung höhergradiger Trendpolynome zur zukünftigen Fortschreibung von Indikatoren grundsätzlich in Frage gestellt werden. Um weitergehende Evaluationen dieses grundsätzlich identifizierten Phänomens zu ermöglichen, wurde auf explizitem Wunsch des Auftraggebers dennoch eine Trendfortschreibung auf Basis der DAS Monitoring Analysen im R-Tool implementiert. Da im Rahmen dieses Vorhabens keine weitergehenden Überarbeitungen bereits vorliegender R-Skripte erfolgen sollten, obliegt es damit derzeit der Sorgfalt der Anwenderinnen und Anwender des Tools, projizierte Trendverläufe auf Basis quadratischer Trendschätzungen eigenständig kritisch zu hinterfragen.<sup>22</sup>

---

<sup>21</sup> Mit  $\epsilon_t$  werden dabei wiederum (wie im gesamten Verlauf dieses Textes) die Ausprägungen eines statistischen Zufallsterms bezeichnet (vgl. hierzu auch Gleichung 1.c). Die Parameter dieser Geradengleichung ( $\gamma_0, \gamma_1$ ) bezeichnen den Ordinaten-Schnittpunkt sowie die Steigung der geschätzten Trendgerade.

<sup>22</sup> Das bisherige R-Skript ließe sich durchaus dahingehend umbauen, dass den Anwenderinnen und Anwendern eine Möglichkeit gewährt würde, lediglich Schätzungen auf Basis linearer Effekte zu extrapolieren. Unter Verweis auf den prototypischen Charakter des im Rahmen dieses Gutachtens entwickelten Tools verbleiben entsprechende Modifikationen aber für zukünftige Entwicklungsaktivitäten.

Die zuvor vorgestellte Idee einer Prüfung der Projektionseigenschaften bei Schätzung erster Differenzen kann in diesem Sinne auch als eine eigene Entwicklung zur weiteren Abschätzung der Sensitivität unterschiedlicher Projektionsansätze interpretiert werden. Um den Anwenderinnen und Anwendern umfassende Möglichkeiten für flexible Vergleiche unterschiedlicher Projektionsverläufe bei Schätzung linearer bzw. quadratischer Trends bieten zu können, wurden im Tool zudem weitere Optionen implementiert. So können die Anwenderinnen und Anwender bspw. (ohne weitere Signifikanz-, bzw. sonstige Fehlspezifikationsprüfungen) für jeden beliebigen Indikator die resultierenden Trendprojektionen bei Anwendung einer linearen, bzw. quadratischen Trendhypothese für individuell auswählbare Beobachtungszeiträume darstellen lassen.

### 3.1.2.2 Herausforderungen bei der Identifikation der Ursachen von Autokorrelation.

#### Generelle Vorbemerkungen

Die im DAS Monitoring Trendbewertungsverfahren angewandte Methodik zur Abbildung von Autokorrelationseffekten kann auf eine lange Tradition in der statistischen Analyse von Zeitreihen verweisen (siehe hierzu bspw. Cochrane und Orcutt 1949). In diesem Sinne ist der Ansatz nicht zu kritisieren. Allerdings ist anzumerken, dass dabei lediglich die mathematische Modellierung eines statistisch beobachteten Symptoms (Autokorrelation) angestrebt wird. Eine weitergehende Identifikation der Ursachen dieses Symptoms wird hingegen nicht angestrebt.

Dieser Aspekt kann kritisiert werden, da viele empirische Zeitreihenanalysen häufig aufgrund ähnlicher Ursachen mit Autokorrelationseffekten konfrontiert werden. Dies bedeutet, dass bei identifizierten Autokorrelationseffekten die „üblichen Verdächtigen“ relativ leicht benannt werden können.<sup>23</sup> Angesichts des üblicherweise relativ kurzen Beobachtungsumfangs der KIS-Indikatoren ist der Versuch einer automatischen Identifikation dieser „üblichen Verdächtigen“ wahrscheinlich wenig zielführend. Stattdessen sollte zukünftig aber erwogen werden, welche ergänzenden statistischen Informationen für eine detailliertere Abschätzung der jeweiligen Autokorrelationsursachen notwendig erscheinen. Auf Basis entsprechender statistischer Informationen könnten die jeweiligen Anwenderinnen und Anwender dann (vor Hintergrund ihrer inhaltlichen Expertise) in die Lage versetzt werden, den der Bewertung zugrundeliegenden Schätzansatz eigenständig (bspw. zur Berücksichtigung von Strukturbrüchen) zu respezifizieren.

#### Strukturbrüche

Grundsätzlich denkbare Spezifikationsalternativen zur Berücksichtigung von Strukturbrüchen sollen am Beispiel von Abbildung 17 exemplarisch verdeutlicht werden. In der oberen Grafik werden dort die historischen Beobachtungen des Primärenergieverbrauchs in Deutschland für den Zeitraum 1990 bis 2016 dargestellt. Bezeichne  $y_t$  die jeweiligen Beobachtungen dieses Indikators zum Zeitpunkt  $t$ , so lässt sich die Hypothese, dass diese Beobachtungen im gesamten Beobachtungszeitraum lediglich durch zufällige Abweichungen von einem linearen Trend erzeugt wurden, wiederum in Form der Gleichung (3) notieren ( $y_t = \gamma_0 + \gamma_1 t + \epsilon_t$ ).<sup>24</sup>

---

<sup>23</sup> Siehe hierzu bspw. auch folgendes Zitat aus der Lehrbuchliteratur: „Häufig liegen der beobachteten Autokorrelation [...] relevante Faktoren wie fehlende Variable, unberücksichtigte Modelldynamik oder Strukturbrüche zugrunde.“ (Winker 2010, S. 181).

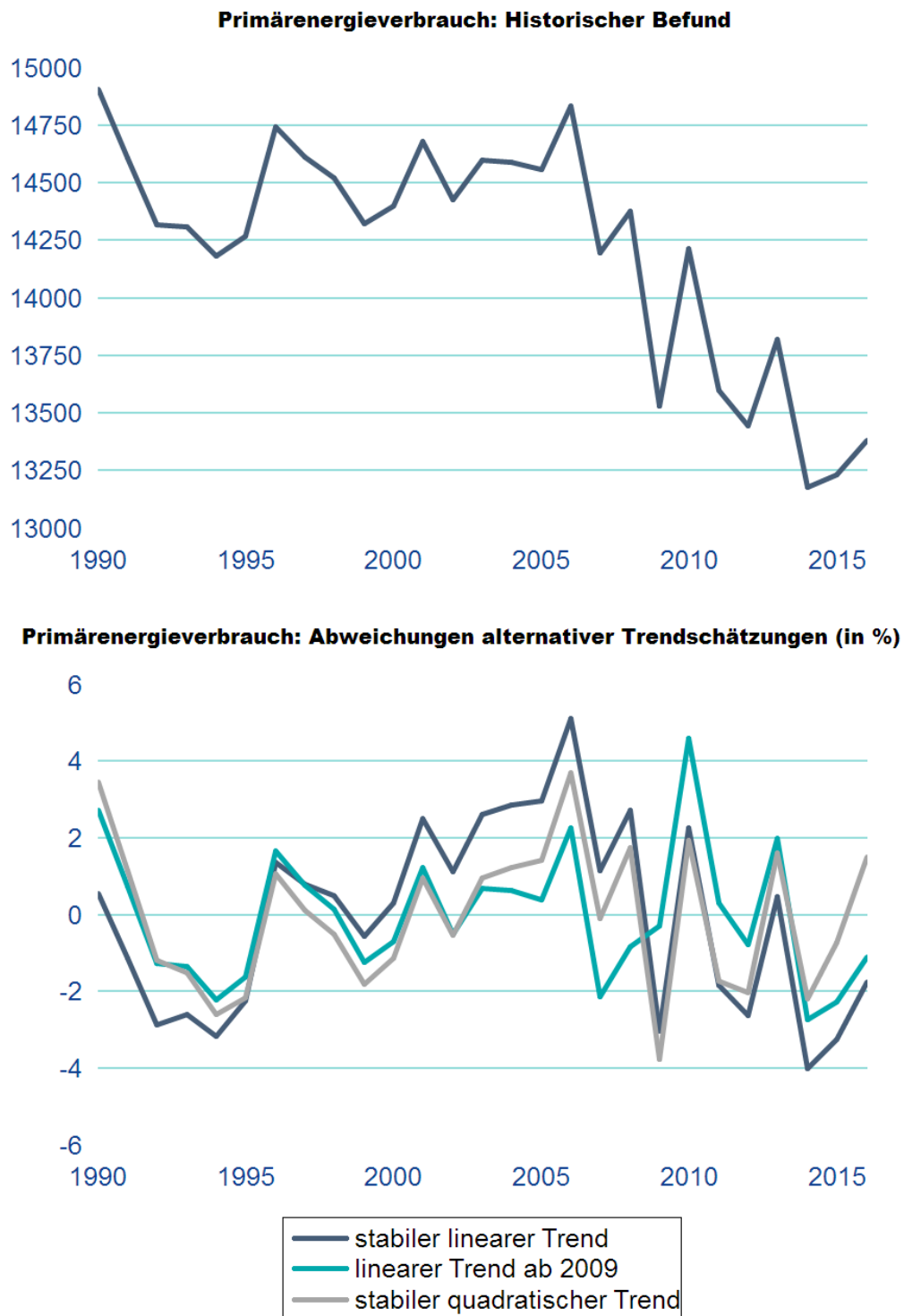
<sup>24</sup> Der Notation in Gleichung (1.c) folgend, werden mit  $\epsilon_t$  wiederum die Ausprägungen eines statistischen Zufallsterms bezeichnet.

Bei einer Kleinst-Quadrate Schätzung der Parameter  $\gamma_0, \gamma_1$  dieser Geradengleichung zeigen sich dann allerdings deutliche Anzeichen auf eine positive Autokorrelation der Residuen.<sup>25</sup> In einfachen Worten ausgedrückt bedeutet dies, dass im Zeitablauf auf positive Abweichungen von der geschätzten Trendgeraden tendenziell weitere positive Abweichungen (bzw. auf negative Abweichungen tendenziell weitere negative Abweichungen) folgen. Dies widerspricht einer zentralen Grundannahme des durch Gleichung (3) beschriebenen Zeitreihenmodells, da zu jedem beobachteten Zeitpunkt die Realisation des Störterms  $\varepsilon_t$ , also dessen Wert, annahmegemäß nicht von den vorherigen Realisationen, also den vorhergehenden Werten des Störterms, abhängig sein sollte.

---

<sup>25</sup> Die Durbin-Watson Statistik, welche auf Autokorrelation erster Ordnung prüft, Werte zwischen Null und Vier annehmen kann und bei Abwesenheit von Autokorrelation erster Ordnung einen Wert in der Nähe von Zwei annehmen sollte, berechnet sich für dieses Beispiel zu 1.05. Auch ein LM-Test zur Prüfung der Nullhypothese „keine Autokorrelation 1. Ordnung der Residuen“ verdeutlicht, dass diese bei einem 5%-Signifikanzniveau deutlich verworfen werden muss.

**Abbildung 17: Empirisches Beispiel zur Frage der Behandlung von Strukturbrüchen am Beispiel Primärenergieverbrauch (in Petajoule)**



Quelle: Eigene Darstellung der Ergebnisse bei Anwendung des R-Skripts zu Schönthaler et al. (2015).

Bei Anwendung des DAS Monitoring R-Skripts wird diese Spezifikation folglich zugunsten einer quadratischen Trendschätzung abgelehnt. Im Rahmen der vorgegebenen

Entscheidungsoptionen ist diese Wahl des Schätzalgorithmus plausibel nachvollziehbar, da der historische Befund durchaus als ein im Zeitablauf nicht-konstantes Trendmuster interpretiert werden kann: Ungefähr bis zum Jahr 2008 scheint der Indikator relativ gleichmäßige Schwankungen um einen konstanten Mittelwert aufzuweisen. Erst danach spricht der visuelle Befund für eine rückläufige Trendentwicklung.

Als alternative Hypothese könnte daher bspw. auch geprüft werden, ob diese Zeitreihe gegenwärtig einem linearen Trendverlauf folgt, welcher allerdings erst im Jahr 2009 einsetzte. Wenn mit  $I_{T^*}$  eine Indikatorvariable bezeichnet wird, welche vor dem Zeitpunkt  $T^*$  jeweils Null ist und danach stets den Wert Eins annimmt, lässt sich diese Hypothese folgendermaßen spezifizieren:

$$(4) y_t = \gamma_0 + \gamma_1 t I_{2009,t} + \epsilon_t.$$

Die untere Grafik der Abbildung 17 zeigt die bei Durchführung der jeweiligen Schätzungen resultierenden Residuen (skaliert als prozentuale Abweichung der historischen Beobachtungen von den jeweils geschätzten Trendverläufen). Die dunkel-graue Kurve korrespondiert dabei zur einfachen linearen Trendschätzung (Gleichung 3). Die grüne Kurve bildet die Residuen bei Schätzung von Gleichung (4) ab, die hell-graue Kurve kennzeichnet die Residuen des vom R-Skript ausgewählten Schätzansatzes

$$(5) y_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \epsilon_t.$$

Ein Vergleich der Schwankungsbreiten dieser Residuenzeitreihen legt nahe, dass sowohl bei Schätzung von Gleichung (4) als auch bei Schätzung von Gleichung (5) eine im Vergleich zur einfachen linearen Trendschätzung (Gleichung 3) bessere historische Anpassung erreicht wird. (Beide Kurven tendieren dazu, im Vergleich zur dunkel-grauen Linie weniger stark von der Nulllinie abzuweichen.) Zudem lässt sich zumindest erahnen, dass beide Schätzungen im Vergleich zur Schätzung eines linearen Trends des gesamten Beobachtungszeitraums keine deutlichen Anzeichen von Autokorrelation aufweisen: Während die historischen Beobachtungen der Jahre 2000 bis 2008 bspw. die geschätzte Trendgerade kontinuierlich übersteigen (dunkel-graue Kurve), weisen die beiden alternativen Residuenzeitreihen keine entsprechend persistenten Abweichungen auf.<sup>26</sup>

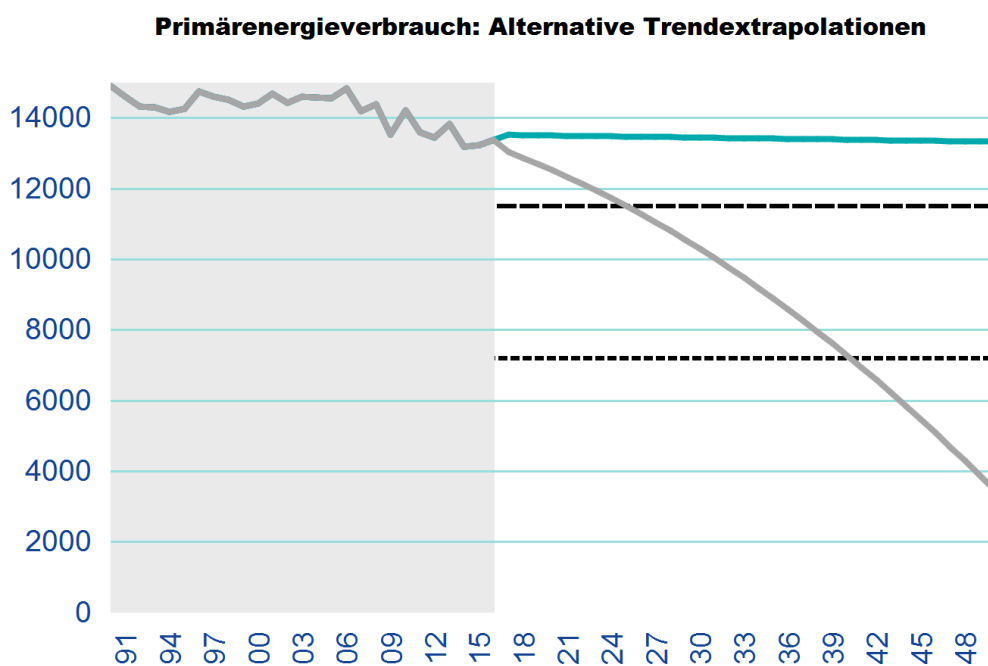
Das betrachtete Beispiel verdeutlicht damit, dass eine Schätzung quadrierter Trendterme gut geeignet ist, Anzeichen für im Beobachtungszeitraum zeitvariante Trendeffekte aufzuspüren. Die Annahme eines langfristig stabilen linearen Trends kann angesichts dieser Befunde sicherlich verworfen werden. Unklar bleibt an dieser Stelle aber zunächst, welche alternative Trendhypothese für zukünftige Trendprojektionen angewandt werden sollte. Unter Berücksichtigung der vorherigen Anmerkungen zur Prognoseeigenschaft von Trendpolynomen, sollte diese Schätzgleichung sicherlich nicht ohne umfassende Prüfung alternativer Spezifikationsansätze für Projektionszwecke angewandt werden.

---

<sup>26</sup> Dass sowohl die grüne wie auch die hell-graue Kurve keine signifikanten Anzeichen von Autokorrelation aufweisen, wird auch durch entsprechende formale Tests (deren Ausweis an dieser Stelle unterlassen wird) bestätigt.

Zumindest belegt Abbildung 18, dass sich die beiden an dieser Stelle exemplarisch miteinander verglichenen Schätzansätze deutlich hinsichtlich ihrer implizierten Prognoseeigenschaften unterscheiden. Die Abbildung zeigt den historischen Verlauf des deutschen Primärenergieverbrauchs (im grau schraffierten Bereich) gemeinsam mit Projektionen auf Basis der Schätzergebnisse für Gleichungen (4) und (5). Die beiden schwarz gestrichelten Linien symbolisieren dabei die jeweiligen Zielwerte gemäß Energiekonzept und Nachhaltigkeitsstrategie der Bundesregierung. Gemäß dieser Zielvorgaben sollte der deutsche Primärenergieverbrauch im Jahr 2020 die obere gestrichelte Linie unterschreiten. Für das Jahr 2050 wird ein Erreichen der unteren gestrichelten Linie angestrebt.

**Abbildung 18: Alternative Trendextrapolationen des Primärenergieverbrauchs (in Petajoule)**



Quelle: Eigene Darstellung der Ergebnisse bei Anwendung des R-Skripts zu Schönthaler et al. (2015).

Kurzfristig stimmen beide Projektionen dahingehend überein, dass der Zielwert des Jahres 2020 bei Aufrechterhaltung des bisherigen Trendverlaufs nicht erreicht wird. Langfristig implizieren beide Projektionen allerdings gravierende qualitative Unterschiede: Die hell-grau dargestellte Projektion des quadratischen Trends unterstellt, dass der Zielwert des Jahres 2050 bereits zu Beginn der 2040er Jahre unterschritten wird. Auf Basis des ab dem Jahr 2009 geschätzten linearen Trends werden hingegen auch langfristig keine nennenswerten weiteren Rückgänge des deutschen Primärenergieverbrauchs erwartet (grüne Kurve).

Derzeit können im Tool keine expliziten Dummy-Regressionen durchgeführt werden. Der hier exemplarisch vorgestellte Vergleich alternativer Trendspezifikationen unter Verwendung einer geeignet definierten Indikator-Variable kann also bislang von den Anwenderinnen und Anwendern des Tools nicht systematisch durchgeführt werden. Dieser Entwicklungsstand beruht auf der Tatsache, dass im Rahmen dieses Vorhabens lediglich ausgewählte bestehende



Verfahren, bzw. Bewertungsverfahren in Anlehnung an entsprechende univariate Bewertungsverfahren implementiert werden sollten. Da das Vorhaben nicht auf umfassende Erweiterungen dieser Verfahren abzielte, verbleibt es zukünftigen Forschungsvorhaben vorbehalten, diesbezügliche Erweiterungsoptionen vertieft methodisch zu diskutieren um diese daraufhin ebenfalls im Tool zu implementieren.

Dabei muss beachtet werden, dass entsprechende Erweiterungen sicherlich nicht als vollständig automatisierte Algorithmen zur datenbasierten Identifikation eines idealen Prognoseansatzes umgesetzt werden können. Bei der individuellen Auswahl einer Strukturbruch-Spezifikation sind die Anwenderinnen und Anwender grundsätzlich angehalten, substantielle Gründe zur Verwendung der ausgewählten Projektionsgleichung zu identifizieren. So unterstellt Gleichung (4) bspw. einen Strukturbruch, indem die Entwicklung des deutschen Primärenergieverbrauchs bis zum Jahr 2008 als zufällige Schwankungen um einen konstanten Erwartungswert interpretiert werden, woraufhin ab dem Jahr 2009 eine langfristig rückläufige Entwicklung einsetzt. Solche Effekte können durch dauerhafte Veränderungen kausaler Einflussfaktoren oder aber auch durch Änderungen in der Art der Datenerfassung hervorgerufen werden.

Im letztgenannten Fall obliegt es grundsätzlich den Anwenderinnen und Anwendern zu erwägen, ob Änderungen in der Art der Datenerfassung eine Verwendung der in Gleichung (4) spezifizierten Trendhypothese inhaltlich tragfähig begründen können.

In der gegenwärtigen Ausbaustufe können mit dem Tool bislang auch keine Anzeichen für dauerhafte Veränderungen kausaler Einflüsse analysiert werden. Hierzu wären statistische Prüfalgorithmen aus dem Bereich der multivariaten Zeitreihenanalysen im Tool zu implementieren. Da dieses Vorhaben aber lediglich auf eine Umsetzung univariater zeitreihenanalytischer Verfahren abzielte, verbleiben auch diesbezügliche Arbeiten als eine Option für zukünftige Toolerweiterungen.

#### Dynamische Spezifikationen

Der vorherige Abschnitt verdeutlichte, dass beobachtete Autokorrelationseffekte durch Strukturbrüche hervorgerufen werden können. Wenn dies der Fall ist, erscheint eine explizite Modellierung der Ursachen dieser Strukturbrüche empfehlenswert. Als für Prognosezwecke eher unbefriedigend erscheint im Vergleich hierzu die vom DAS Monitoring Verfahren stattdessen vorgenommene statistische Modellierung dieser Autokorrelationsstrukturen.

Dies bedeutet allerdings nicht, dass Trendschätzungen mit autokorrelierten Residuenstrukturen für Prognosezwecke generell ungeeignet sind: Wenn die beobachteten Autokorrelationseffekte nicht auf Strukturbrüche sondern auf dynamische Anpassungsprozesse zurückzuführen sind, kann die Fortschreibung eines linearen DAS-Monitoring Trends mit autokorrelierten Störtermen bspw. durchaus gerechtfertigt werden. Abbildung 19 verdeutlicht diese Überlegungen anhand des folgenden Beispiels: Die durchgezogene dunkelgraue Linie zeigt den Verlauf einer simulierte Indikatorreihe im Zeitraum 1991 bis 2015. In der Simulation wurde dabei unterstellt, dass diese Zeitreihe einem stabilen linearen Trend folgt, welcher langfristig einen Rückgang dieser Zeitreihe um ca. 1,1% pro Jahr vorgibt. Bezeichne  $\tau_t$  den natürlichen Logarithmus des unterstellten Trendverlaufs, so lässt sich die in dieser Simulation unterstellte Trendspezifikation dabei näherungsweise folgendermaßen zusammenfassen:  $\tau_t = 7,062 - 0.011 \cdot t$ .<sup>27</sup>

---

<sup>27</sup> Die Festlegung der in diesem Beispiel angewandten Simulationsparameter erfolgte in Anlehnung an eigene Befunde bei Analyse des Indikators „Emission der von der UN-Klimarahmenkonvention abgedeckten Treibhausgase“.



In der Simulation wurde nun aber nicht unterstellt, dass sämtliche beobachtbaren Abweichungen von diesem Trendverlauf lediglich auf zufällige Störterme zurückzuführen sind. Stattdessen wurde eine sogenannte Fehler-Korrektur Spezifikation angewandt, welche einen dynamischen Anpassungsprozess unterstellt. Dies bedeutet, dass einmalige Abweichungen des Indikators vom unterstellten Trendverlauf erst im Verlauf mehrerer Folgeperioden abgebaut werden.

Bezeichne  $y_t$  die logarithmierten Werte der simulierten Indikatorreihe, so lässt sich der simulierte Anpassungsprozess unter Betrachtung der ersten Differenzen dieser Variable folgendermaßen notieren:

$$\Delta y_t = -0,75 \cdot (y_{t-1} - \tau_t) + \varepsilon_t, \varepsilon_t \sim \text{iid } N(0, \sigma^2).$$

Es wurde also unterstellt, dass der Indikator die Abweichung seines Vorjahreswertes  $y_{t-1}$  vom langfristigen Trendniveau  $\tau_t$  binnen eines Jahres lediglich zu 75% reduzieren kann, wobei dieser Anpassungsprozess durch zufällige Störeinflüsse  $\varepsilon_t$  überlagert wird.<sup>28</sup> Sämtliche Parameter dieses Fehler-Korrektur-Ansatzes können dann unter Anwendung der üblichen Methode der Kleinsten-Quadrate anhand folgender Schätzgleichung bestimmt werden:

$$(6) \Delta y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot t + \varepsilon_t.$$

Der anhand dieser Schätzung für die simulierte Indikatorreihe identifizierte lineare Trendverlauf ist in Abbildung 19 als grün gestrichelte Trendlinie dargestellt. Während an dieser Stelle aus Platzgründen auf eine detaillierte Repräsentation numerischer Schätzergebnisse sowie weiterführende Anmerkungen zur Interpretation der Parameter von Gleichung (6) verzichtet wird,<sup>29</sup> ist die Tatsache bemerkenswert, dass diese Trendlinie kaum von der zweiten, in dieser Abbildung hellgrau markierten Trendschätzung unterschieden werden kann.

Die hellgraue Linie repräsentiert die Schätzergebnisse bei Anwendung der DAS Monitoring Routinen auf die (natürlichen Logarithmen) der simulierten Indikatorzeitreihe. Gemäß der zuvor gewählten Notation lassen sich diese Ergebnisse (gerundet) folgendermaßen zusammenfassen:

$$y_t = 7,066 - 0,011 \cdot t + u_t, \quad u_t = 0,31 \cdot u_{t-1} + \varepsilon_t.$$

Unter zusätzlicher Schätzung eines autoregressiven Prozesses erster Ordnung für die Störterme  $u_t$  wird damit der tatsächlich simulierte fallende lineare Trend nahezu perfekt abgeschätzt. Da die identifizierten Autokorrelationsstrukturen nicht auf sonstige Fehlspezifikationen zurückzuführen sind, sondern lediglich dem unterstellten dynamischen Anpassungsprozess des Indikators geschuldet sind, ist in diesem Beispiel eine Fortschreibung der bisherigen Indikatorentwicklung auf Basis der DAS-Monitoring offensichtlich gerechtfertigt.

Um den Anwenderinnen und Anwendern des Tools eine Möglichkeit zur Durchführung eigener entsprechender Abwägungen bei der Betrachtung von DAS Monitoring Ergebnissen zu bieten,

---

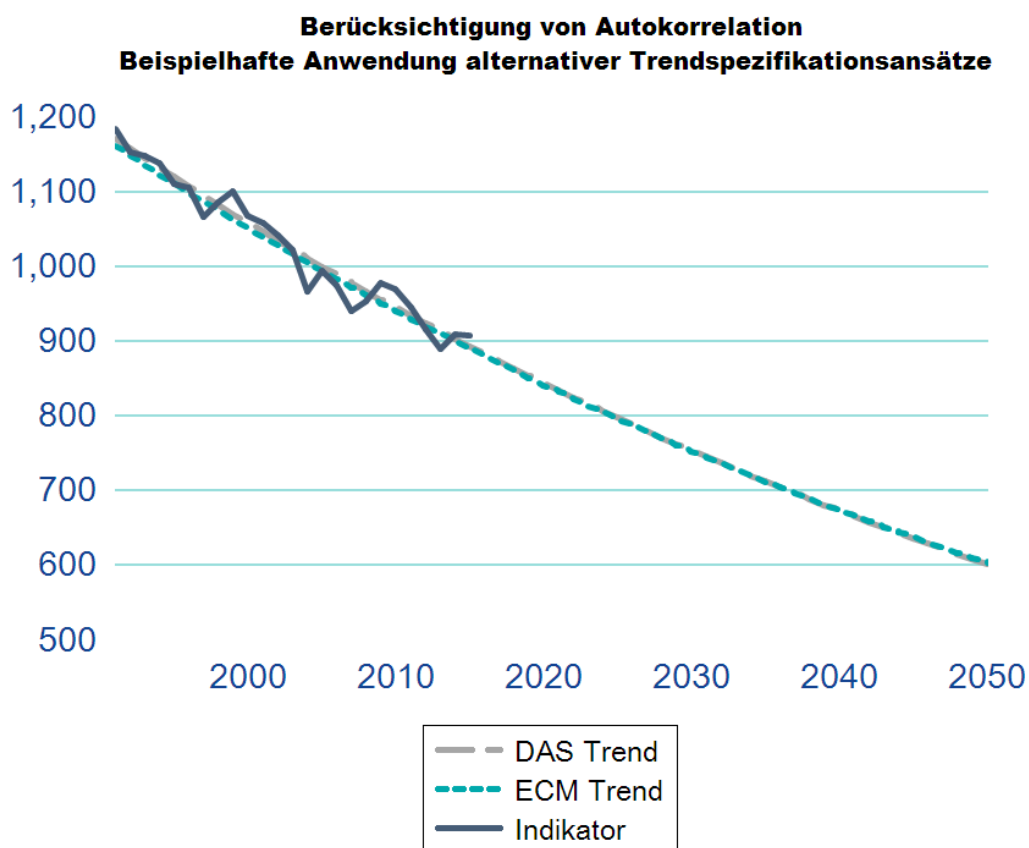
<sup>28</sup> Für die Standardabweichung des simulierten Störterms  $\varepsilon$  wurde dabei ein Wert von (näherungsweise) 0,017 unterstellt.

<sup>29</sup> Erläuternde Anmerkungen zur Schätzung von Fehler-Korrektur Modellen sowie der Interpretation der resultierenden Parameterwerte können bspw. Winker (2010, S. 259-265) entnommen werden.

prüft der implementierte eigene prototypische Bewertungsalgorithmus automatisch die Anpassungsgüter von Fehler-Korrektur Spezifikationen gemäß Gleichung (6). Eine kurze Vorstellung der dabei verfolgten Vorgehensweise findet sich in Kapitel 3.2.3.

Wenn, wie im obigen Beispiel beobachtet, die Projektionseigenschaften einer entsprechenden Schätzung kaum von den Projektionseigenschaften auf Basis der DAS Monitoring Befunde abweicht, können diese Befunde insgesamt als ausgesprochen robust angesehen werden. Im Gegensatz hierzu sollten deutliche Abweichungen zwischen den jeweils prognostizierten Indikatorverläufen von den Anwenderinnen und Anwendern als mahnendes Signal zur intensivierten inhaltliche Prüfung der jeweiligen Bewertungsergebnisse aufgefasst werden. Die konkrete Festlegung der dabei anzuwendenden weiterführenden Prüfroutinen bleibt zukünftigen methodischen Forschungsaktivitäten überlassen.

**Abbildung 19: Trendextrapolationen unter Verwendung alternativer dynamischer Modellspezifikationen**



Quelle: Eigene Berechnungen sowie eigene Darstellung der Ergebnisse bei Anwendung des R-Skripts zu Schönthaler et al. (2015).

### 3.1.2.3 Beachtung des Scheinkorrelationsrisikos

Methodischer Hintergrund: Erläuterung des Phänomens stochastischer Trendmuster

Als vordringlich zu adressierender Kritikpunkt wurde im Rahmen dieses Gutachtens die Tatsache identifiziert, dass der Trend-Identifikationsalgorithmus in der Tradition von Martone

(2013) auch in der weiteren Verwendung durch Schönthaler et al. (2015) sämtliche statistischen Beobachtungen ausschließlich dazu nutzt, Anzeichen für das Vorliegen deterministischer Trendphänomene aufzuspüren. In Ermangelung einer stochastischen Prüfung alternativer Erklärungsansätze tendiert dieses Verfahren dabei dazu, zu häufig signifikante Trendsignale anzuzeigen.

Um diesen Kritikpunkt anschaulich darstellen zu können sei zunächst daran erinnert, dass die beobachtbaren Entwicklungen der Umweltindikatoren selbstverständlich nicht durch die Zeit verursacht werden. Vielmehr resultieren diese Befunde aus dem komplexen Zusammenspiel unterschiedlicher und häufig divergierender kausaler Einflussfaktoren.<sup>30</sup> Bei der ausschließlichen Analyse einer Indikatorzeitreihe auf Anzeichen langfristiger Trendentwicklungen wird allerdings auf eine explizite Berücksichtigung entsprechender Kausaleffekte verzichtet. Dies ist dadurch motiviert, dass die statistische Identifikation dieser Kausaleffekte in der Regel zeitaufwendig ist. Wenn aber keine Evaluation der für die zeitliche Entwicklung eines Indikators ursächlichen Kausalitäten erfolgt, kann zunächst nicht ausgeschlossen werden, dass die vorliegenden Beobachtungen lediglich einen zufälligen Befund repräsentieren.

Beispielhaft sei hierzu auf Abbildung 20 verwiesen. Die dort abgebildeten Zeitreihen zeigen drei unterschiedliche Realisationen eines datenerzeugenden Prozesses, bei welchem die dynamische Entwicklung der Zeitreihe lediglich durch zufällige jährliche Veränderungen verursacht wurde.

Formal lässt sich diese Zeitreihenmodellierung folgendermaßen zusammenfassen:

$$(7) \quad y_t = y_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{iid } N(0, \sigma^2).$$

Simuliert wurden Zeitreihen mit einem Umfang von 200 Beobachtungen, wobei für die Standardabweichung ein numerischer Wert von fünf vorgegeben wurde.

Offensichtlich sind die für diese Abbildung ausgewählten Zeitreihenrealisationen durch langfristig ansteigende Trendmuster gekennzeichnet. Wichtig ist nun aber die Erkenntnis, dass die beste Prognose für Zeitreihen, welche dem durch Gleichung (7) beschriebenen datenerzeugenden Prozess folgen (ein sogenannter „Random Walk“), durch den letzten beobachteten Wert dieser Zeitreihe gegeben ist: Der durch Abbildung 20 vermittelte visuelle Eindruck langfristig positiver Trendverläufe beruht lediglich auf einer zufälligen Dominanz der positiven Realisationen des Störterms  $\varepsilon_t$  im Simulationszeitraum. Sämtliche zukünftige Entwicklungen werden wiederum ausschließlich durch zufällige Realisationen des Störterms bestimmt. Da diese voneinander unabhängig sind, besteht kein Grund zu der Annahme, dass auch zukünftig die positiven Realisationen die weitere Entwicklung dominieren werden.

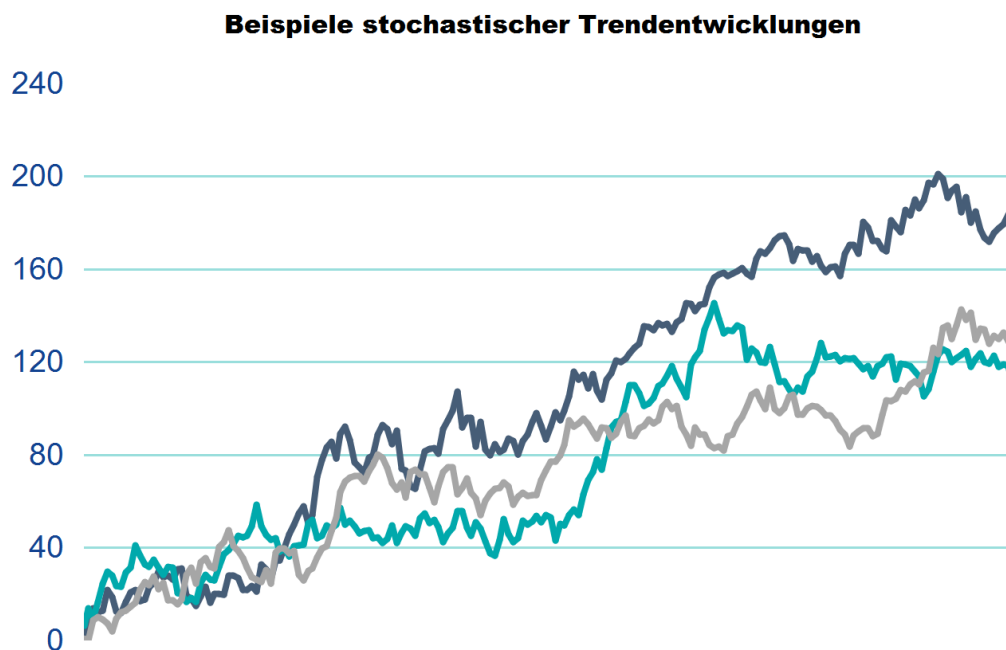
Die soeben beispielhaft vorgestellten Trendverläufe werden in der Fachliteratur als „stochastische Trends“ bezeichnet. Im Gegensatz hierzu werden zeitabhängige Trendprozesse, wie sie bspw. durch Gleichung (3) beschrieben werden, als „deterministische Trends“ bezeichnet. Angesichts der vollkommen unterschiedlichen Projektionseigenschaften von

---

<sup>30</sup> Kausale Einflussfaktoren mögen bspw. (u.a.) wirtschaftliche Triebkräfte (wie Rohstoffpreise, Investitionsdynamik oder die Entwicklung globaler Handelsstrukturen im Fall der Primärrohstoffnutzung für inländischen Konsum und Investitionen) und individuelle Präferenzen (bspw. in Form der persönlichen Wertschätzung von Umwelt- und Naturschutzmaßnahmen als Triebfeder eines persönlichen Engagements im Umwelt- und Naturschutz) im Zusammenspiel mit hierauf einwirkenden legislativen Initiativen sein. Entsprechende Einflussfaktoren determinieren für sämtliche Indikatoren jeweils eine individuelle Entwicklungsdynamik, welche dann (ggf.) in einer stabilen zeitlichen Korrelation resultiert.

stochastischen und deterministischen Trends wäre es offensichtlich grob irreführend, wenn für stochastische Trendrealisationen deterministische Trendverläufe unterstellt würden. Da der in Schönthaler et al. (2015) beschriebene Algorithmus allerdings nicht zu Projektionszwecken entwickelt wurde, wird dieser Aspekt dort bislang vollständig ausgeblendet.

**Abbildung 20: Stochastische Trends als Realisation eines simulierten Random Walks**



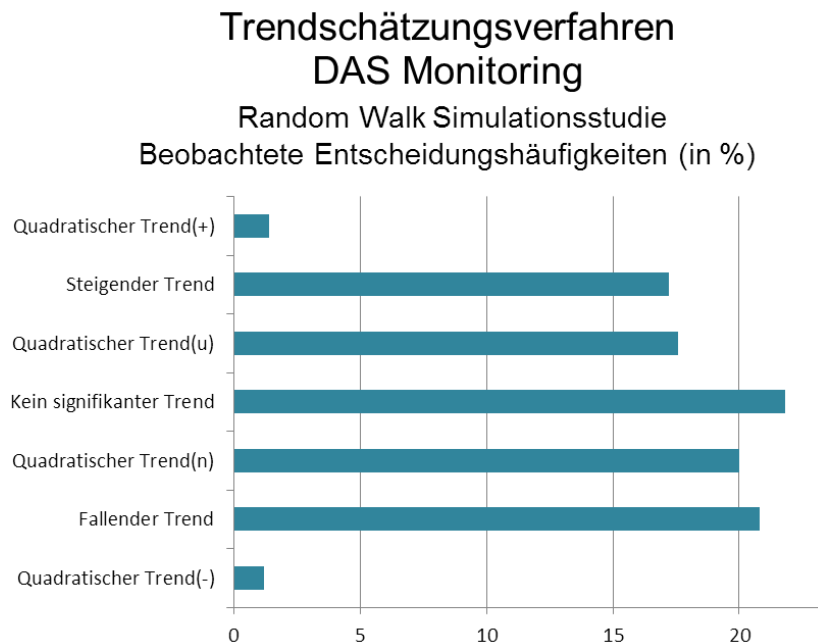
Quelle: Eigene Simulationsergebnisse. Auswahl Random-Walk Realisationen mit positivem Muster

#### Empirischer Befund: Ergebnisse einer eigenen Monte Carlo Studie

Die statistischen Implikationen bei Anwendung üblicher Regressionsansätze auf stochastische Trendzeitreihen sind im Forschungsfeld der Ökonometrie grundsätzlich hinlänglich bekannt. Wie bspw. Granger und Newbold (1974) dokumentiert haben, muss daher auch bei Anwendungen der DAS Monitoring Routinen auf stochastische Trendreihen von der Existenz sogenannter Scheinkorrelations-Risiken ausgegangen werden. Das Ausmaß entsprechender Verzerrungen kann für den gegebenen Algorithmus allerdings nicht im Vorhinein pauschal abgeschätzt werden.

Im Rahmen dieses Sachverständigenutachtens wurde daher eine eigene Monte Carlo Studie zur numerischen Abschätzung des implizierten Scheinkorrelations-Risikos des vorliegenden R-Skripts durchgeführt. Auf Basis des durch Gleichung (7) beschriebenen datenerzeugenden Prozesses wurden hierzu 1000 Zeitreihen mit einem Umfang von jeweils 200 Beobachtungen generiert. Diese wurden daraufhin mit dem am UBA vorliegenden R-Skript auf Anzeichen deterministischer Trends analysiert.

**Abbildung 21: Abschätzung des Scheinkorrelationsrisikos**



Quelle: Eigene Ergebnisse unter Anwendung des in Schönthaler et al. (2015) beschriebenen Schätzalgorithmus.

Der zusammenfassende Befund dieses Simulationsexperiments kann Abbildung 19 entnommen werden: Das R-Skript entschied in weniger als 22% aller Fälle richtigerweise, dass kein signifikanter deterministischer Trend vorlag. Dies bedeutet, dass für annähernd vier von fünf Random Walk-Realisationen irrtümlich signifikante Trendanzeigen identifiziert wurden. Formaler ausgedrückt: Wenn eine Zeitreihe stochastische Trendeigenschaften aufweist, liegt das empirische Signifikanzniveau der Routinen von Schönthaler et al. nicht (wie angestrebt) bei 5% sondern (zumindest in diesem Simulations-Setting) bei (annähernd) 80%.

Dieses Simulationsbeispiel mag zunächst artifiziell anmuten. Trotzdem muss festgehalten werden, dass damit eine ernsthafte empirische Herausforderung für die hier evaluierte, derzeit im Rahmen des DAS Monitoring gängige Form der Indikatorenbewertung sichtbar wird.

## 3.2 Methodische Weiterentwicklungen im Zuge der Toolentwicklung

### 3.2.1 Test des Integrationsgrades

Gemäß persönlicher Einschätzung repräsentieren die soeben aufgezeigten Scheinkorrelationsrisiken zentrale methodische Herausforderungen, welche bei einer Verwendung der DAS Monitoring Schätzungen zu Prognosezwecken grundsätzlich zu beachten sind. Um einen diesbezüglich beschreibbaren Weg zur Reduktion dieser Risiken zumindest explorativ vorstellen zu können, zeichnet sich das zum Projektabschluss an das Umweltbundesamt übergebene Tool auch durch einen eigenständigen (derzeit sicherlich noch als prototypisch einzuordnenden) Bewertungsalgorithmus aus, welcher geeignet scheint, dass soeben vorgestellte Scheinkorrelationsrisiko deutlich zu minimieren.

### 3.2.1.1 Theoretische Anmerkungen: Herleitung eines Einheitswurzeltests im Kontext der DAS Monitoring Routinen

Zur Prüfung der Hypothese, ob eine betrachtete Zeitreihe stochastische Trendeigenschaften aufweist, wurden in der statistischen Literatur im Verlauf der vergangenen Jahrzehnte diverse sogenannten „Unit Root Tests“ (deutsch: Einheitswurzeltests) vorgeschlagen. Die Grundidee des wohl bekanntesten Testansatzes wurde dabei von Dickey und Fuller (1979) vorgestellt. An dieser Stelle soll kurz methodisch erläutert werden, warum im Kontext der DAS Monitoring Routinen die Verwendung dieses Testansatzes formal naheliegend erscheint. Als Beispiel soll hierzu ein linearer Trend betrachtet werden, wobei die jeweiligen Abweichungen des Indikators von diesem Trendverlauf einem autoregressiven Prozess erster Ordnung folgen. Gemäß der bisherigen Notationen wird also die Gültigkeit des folgenden Prozesses unterstellt:

$$y_t = \gamma_0 + \gamma_1 t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, \sigma^2).$$

Diese Notation lässt sich dann offensichtlich folgendermaßen umformen:

$$\begin{aligned} y_t &= \gamma_0 + \gamma_1 t + \rho(y_{t-1} - \gamma_0 - \gamma_1(t-1)) + \epsilon_t \\ &= (1 - \rho)\gamma_0 + \gamma_1 t - \rho\gamma_1(t-1) + \rho y_{t-1} + \epsilon_t \\ &= (1 - \rho)\gamma_0 + \gamma_1(t - \rho(t-1)) + \rho y_{t-1} + \epsilon_t \\ (8) \quad &= (1 - \rho)\gamma_0 + \gamma_1 \rho + (1 - \rho)\gamma_1 t + \rho y_{t-1} + \epsilon_t. \end{aligned}$$

Diese Umformungen verdeutlichen für das gewählte Beispiel, dass die in den DAS Routinen angewandten iterativen Maximum Likelihood Schätzer durch übliche Kleinst-Quadrate Schätzungen einer Regressionsgleichung folgender Form ersetzt werden könnten:  $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 t + \epsilon_t$ .<sup>31</sup> Da vorläufige eigene Berechnungen darauf schließen lassen, dass keine nennenswerten numerischen Ergebnisvariationen bei Anwendung entsprechender Kleinst-Quadrate Ansätze anstelle der in den DAS Routinen implementierten Maximum Likelihood Verfahren zu erwarten sind, wird allerdings auf eine weitergehende Diskussion dieses methodischen Aspekts im Rahmen dieses Gutachtens verzichtet.

Im Zusammenhang mit der an dieser Stelle betrachteten Scheinkorrelationsthematik ist die Betrachtung von Gleichung (8) hilfreich um darauf hinzuweisen, dass die hier unterstellte lineare Trendhypothese nur dann sinnvoll definiert ist, wenn der Autokorrelationsparameter  $\rho$  kleiner als Eins ist. Im Spezialfall  $\rho = 1$  beschreibt Gleichung (8) hingegen einen Random Walk Prozess, da dann offensichtlich gilt:

$$(9) \quad y_t = \gamma_1 + y_{t-1} + \epsilon_t.^{32}$$

Im Umkehrschluss folgt hieraus, dass im Kontext der DAS Monitoring Routinen das Scheinkorrelationsrisiko durch einen Test der Hypothese  $H_0: \rho = 1$  gegen die Alternativhypothese  $H_1: \rho < 1$  abgeschätzt werden kann. Wenn diese Nullhypothese für gegebene Indikatorbeobachtungen durch statistische Testverfahren verworfen wird, ist davon auszugehen, dass die betrachtete Indikatorzeitreihe keine Random Walk Eigenschaften aufweist.

<sup>31</sup> Die strukturellen Parameter  $\rho, \gamma_0$  und  $\gamma_1$  können dann unter Beachtung folgender Restriktionen konsistent aus den beobachteten Schätzergebnissen hergeleitet werden:  $\rho = \beta_1, \gamma_1 = \frac{\beta_2}{1-\beta_1}, \gamma_0 = \frac{\beta_0}{1-\beta_1} - \frac{\beta_1 \beta_2}{(1-\beta_1)^2}$ .

<sup>32</sup> Im Vergleich zu der in Gleichung (7) erfolgten Vorstellung eines einfachen Random Walk Prozesses (welcher aus dieser Darstellung im Spezialfall  $\gamma_1 = 0$  resultieren würde), wird ein allgemeinerer Random Walk der Form  $y_t = \gamma_1 + y_{t-1} + \epsilon_t$  gemeinhin als Random Walk mit Drift bezeichnet.

Da damit das zuvor vorgestellte Scheinkorrelationsrisiko ausgeschlossen wird, können die Schätzergebnisse der DAS Monitoring Routinen dann wie gewohnt interpretiert werden.

Der von Dickey und Fuller (1979) vorgestellte Ansatz zum Test dieser Hypothese beruht auf einer einfachen Regressionsgleichung. Im Fall des durch Gleichung (9) beschriebenen Random Walks mit Drift kann zur Herleitung der von Dickey und Fuller vorgeschlagenen Testgleichung von beiden Seiten dieser Gleichung  $y_{t-1}$  abgezogen werden. Da sich dieser Ausdruck dann zu  $y_t - y_{t-1} = \Delta y_t = \gamma_1 + \varepsilon_t$  reduziert, basiert der sogenannte Dickey Fuller Test in diesem Fall auf einer Schätzung der folgenden Regressionsgleichung:<sup>33</sup>

$$(10) \quad \Delta y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t.$$

Die im Fall des Random Walks für Gleichung (9) geltende Hypothese  $H_0: \rho = 1$  kann dann auf Basis der Schätzergebnisse für Gleichung (10) anhand der von Regressionspaketen üblicherweise ausgewiesenen t-Statistik des Parameters  $\beta_1 = \rho - 1$  geprüft werden.<sup>34</sup>

### 3.2.1.2 Im Tool implementierter Testansatz

Eine eigene Anwendung dieses sogenannten Augmented Dickey Fuller (ADF-)Tests auf die in Kapitel 3.1.2.3 analysierten Random Walk Zeitreihen verdeutlicht, dass das Problem des schwachen empirischen Signifikanzniveaus bei Analyse stochastischer Trends hierdurch deutlich reduziert werden kann: Wenn für sämtliche der den Befunden der Abbildung 21 zugrundeliegenden Random Walk-Realisationen jeweils ein ADF-Test (unter Annahme eines deterministischen Trends sowie unter Verwendung der von MacKinnon (1996) dokumentierten kritischen Werte) durchgeführt wird, stimmen theoretisches sowie empirisches Signifikanzniveau (im Rahmen üblicher statistische Schwankungen) miteinander überein. So wird bspw. bei einem vorgegebenen Signifikanzniveau von 10% lediglich in 10.1% aller simulierten Fälle fälschlicherweise entschieden, dass die vorliegende Zeitreihe keine Realisation eines Random Walks repräsentiert.

Auf Basis dieses Befundes wurde in dem an das UBA übergebene Tool auch ein vorläufiger Trend-Identifikationsalgorithmus implementiert, welcher die Ergebnisse des DAS Monitoring-Verfahrens zur Wahl der Spezifikation der angewandten ADF-Testgleichung nutzt. Die dabei angewandten Entscheidungsregeln können folgender Tabelle entnommen werden.<sup>35</sup>

<sup>33</sup> Um die Darstellung zu vereinfachen wird hier nur der grundsätzliche Ansatz des Dickey Fuller Tests erläutert. Im Tool wird automatisch eine sogenannte Augmented Dickey Fuller (ADF) Testroutine angewandt. Im Vergleich zu Gleichung (10) werden dabei verzögerte Terme der abhängigen Variable dieser Testgleichung als weitere Regressoren mit in die Regressorliste aufgenommen, um eventuelle Autokorrelationseffekte berücksichtigen zu können. Weitergehende Details der angewandten Testprozedur können auch der Dokumentation des R-Pakets „CADFtest“ entnommen werden, welches zur Durchführung der ADF-Tests im Tool angewandt wird.

<sup>34</sup> Unter Gültigkeit der Nullhypothese können zur Beurteilung der statistischen Signifikanz der so berechneten t-Statistik nicht die üblicherweise tabellierten Werte herangezogen werden. Die in diesem Fall benötigten kritischen Werte werden im Tool durch das implementierte R-Paket „CADFtest“ zur Verfügung gestellt. Anwenderinnen und Anwender des Tools müssen daher lediglich das anzuwendende Signifikanzniveau bei Durchführung dieser Tests vorgeben, um eine automatisierte Auswertung dieser Testroutinen anhand der unter der Nullhypothese anzuwendenden kritischen Werte zu erhalten.

<sup>35</sup> Während also, wie im vorherigen Beispiel ausführlich hergeleitet, im Fall eines durch die DAS Monitoring Routinen identifizierten linearen Trends der ADF-Test auf Schätzung eines Regressionsansatzes in Form der Gleichung (10) basiert, wird diese Spezifikation im Fall eines quadratischen Trends oder aber eines nicht signifikanten Trends wie in Tabelle 2 angegeben variiert.



**Tabelle 2: Entscheidungsregeln bei Anwendung der ADF-Tests**

DAS Monitoring Befund	(Grund-) Spezifikation der ADF-Testgleichung
Kein signifikanter Trend	$\Delta y_t = \beta_1 \cdot y_{t-1} + \varepsilon_t$
Linearer Trend	$\Delta y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \varepsilon_t$
Quadratischer Trend	$\Delta y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot t + \varepsilon_t$

Auf Basis der nach Vorgabe der DAS Monitoring Befunde ausgewählten ADF-Testgleichung wird daraufhin vom Tool in einem mehrstufigen automatisierten Verfahren der (sogenannte) Integrationsgrad der analysierten Zeitreihe anhand der t-Statistik des geschätzten Parameters  $\beta_1$  identifiziert.<sup>36</sup> Inhaltlich können anhand dieser Testergebnisse folgende Fälle unterschieden werden:

- ▶ I(0), die Zeitreihe ist „(Trend-)Stationär“: Die DAS Monitoring Ergebnisse können als robuste empirische Befunde angesehen werden.
- ▶ I(1), die ersten Differenzen der Zeitreihe sind „(Trend-)Stationär“: Die DAS Monitoring Ergebnisse sollten nicht unmittelbar zur Trendprojektion verwendet werden.
- ▶ I(2), die zweifach differenzierte Zeitreihe ist „(Trend-)Stationär“: Die DAS Monitoring Ergebnisse sollten nicht unmittelbar zur Trendprojektion verwendet werden.

### 3.2.2 Trendprojektionen unter Berücksichtigung des Integrationsgrades der Indikatoren

Wenn ein analysierter Indikator als I(1)- bzw. I(2)-Zeitreihe klassifiziert wird, muss davon ausgegangen werden, dass die in den DAS Monitoring Routinen angewandten Teststatistiken nicht ihren üblichen Verteilungen folgen. Dies bedeutet, dass die DAS Routinen dann fälschlicherweise zu häufig signifikante Trend-Befunde anzeigen. Damit stellt sich die Frage nach alternativen Schätzansätzen, welche im I(1)- und I(2)-Fall angewandt werden können, um die DAS-Befunde inhaltlich weiter zu evaluieren. Im Tool wurde hierzu ein eigener Schätzalgorithmus implementiert, welche auf der soeben vorgestellten ADF-Prozedur basiert und an dieser Stelle wiederum für das Beispiel der Schätzung eines linearen Trends kurz vorgestellt werden soll.

Angenommen, die dynamischen Eigenschaften eines vorliegenden Indikators können durch einen einfachen deterministischen Trend angemessen abgebildet werden (es gilt also das durch Gleichung 2 beschriebene Zeitreihenmodell). Dann gilt:

$$\begin{aligned} \Delta y_t &= y_t - y_{t-1} = \gamma_0 + \gamma_1 t + \epsilon_t - (\gamma_0 + \gamma_1(t-1) + \epsilon_{t-1}) \\ &= \gamma_1 + \epsilon_t - \epsilon_{t-1}. \end{aligned}$$

Aufgrund der Verteilungsannahmen des Störterms  $\epsilon_t$  ist dann folglich garantiert, dass bei Schätzung der Regressionsgleichung

<sup>36</sup> Die Festlegung der Anzahl der dabei (zur Vermeidung autokorrelierter Störterme der Testgleichung) berücksichtigenden verzögerten Differenzen erfolgt nach Maßgabe des sogenannten „Akaike-Informationskriteriums“ (Akaike 1974).



$$(11) \quad \Delta y_t = \hat{\beta}_0 + \hat{\varepsilon}_t$$

die Hypothese  $H_0: \beta_0 = 0$  anhand der üblicherweise tabellierten kritischen Werte geprüft werden kann. Folglich kann anhand des Schätzwertes  $\hat{\beta}_0$  aus Gleichung (11) die Existenz eines deterministischen linearen Trends geprüft werden.

Die vorherige Herleitung des ADF-Ansatzes verdeutlicht zudem, dass diese Schätzgleichung auch bei Vorliegen eines Random Walks mit Drift als Projektionsansatz geeignet ist. Denn in diesem Fall folgt aus Gleichung (9), dass die zukünftigen Veränderungen des Indikators  $\Delta y_t$  durch den deterministischen Drift-Parameter  $\gamma_1$  sowie zufällige Störterme beeinflusst werden. Da dieser Drift-Parameter anhand von Gleichung (11) unverzerrt geschätzt und seine Signifikanz dabei unter Gültigkeit der üblichen statistischen Verteilungsannahmen verlässlich bestimmt werden kann, erscheinen Fortschreibungen auf Basis einer Schätzung der Gleichung (11) sowohl als geeigneter Ansatz zur Projektion linearer deterministischer Trends wie auch zur Abschätzung der zukünftigen Entwicklung von Random Walks mit Drifts. Die Existenz einfacher Random Walk Prozesse kann anhand dieser Schätzung ebenfalls leicht überprüft werden. Wenn bei Schätzung von Gleichung (11) die Hypothese  $H_0: \beta_0 = 0$  nicht verworfen werden kann, ist von der Existenz eines einfachen Random Walk Prozesses auszugehen.<sup>37</sup> In diesem Fall ist die beste Prognose für zukünftige Indikatorentwicklungen durch den letzten beobachteten Wert gegeben.

Vor Hintergrund dieser Überlegungen wurde im Tool ein eigenständiger Trendidentifikations- und -projektionsalgorithmus implementiert, welcher für zuvor als I(1) oder I(2) identifizierte Zeitreihen eine Fortschreibung auf Basis der in Tabelle 2 aufgeführten jeweiligen Testgleichungen vornimmt. Die jeweiligen Test-Spezifikationen (welche für eventuelle Autokorrelationseffekte ggf. durch den Einschluss mehrerer verzögerter Realisationen der abhängigen Variablen auf der Regressorliste der Testgleichung kontrollieren) werden dabei zunächst auf Signifikanz sämtlicher erklärender Variable getestet und ggf. unter Ausschluss nicht signifikanter Regressoren neu geschätzt. Die implizierten Ergebnisse der so re-spezifizierten Testgleichung werden dann als alternative Projektion des zukünftigen Verlaufs des analysierten Indikators in ihrem Niveau abgebildet, d.h. im Tool erfolgt automatisch eine Projektion von  $\hat{y}_{t+1} = y_t + \Delta \hat{y}_t$ , wobei  $\hat{y}$  die Niveau-Projektion der Zeitreihe  $y$  bezeichnet.

Die so generierte Trendprognose beginnt damit stets im letzten historischen Beobachtungswert. Damit sind diese Fortschreibungen unmittelbar mit jenen in Anlehnung an die Verfahren von Eurostat sowie dem Statistischen Bundesamt vergleichbar, da diese ebenfalls stets den letzten historischen Beobachtungswert exakt reproduzieren. Im Vergleich zu Fortschreibungen auf Basis der DAS-Identifikationsroutinen wird hierbei allerdings ein grundsätzlicher Unterschied erkennbar: Optional kann im Tool auch bei den DAS-Projektionen eine Angleichung des Startwerts der Projektion an den letzten historisch beobachteten Indikatorwert vorgenommen werden. Üblicherweise wird durch diese Routinen allerdings ein Projektionspfad generiert, welcher den letzten historischen Beobachtungswert nicht exakt reproduziert. – Während sich beide Vorgehensweisen inhaltlich begründen lassen, sollte bei zukünftigen Entwicklungen standardisierter Projektionsverfahren diesbezüglich eine einheitliche Vorgehensweise festgelegt werden.

<sup>37</sup> Dieser Zusammenhang ist dadurch begründet, dass gemäß Gleichung (5) die ersten Differenzen eines einfachen Random Walks Realisationen eines identisch normalverteilten Fehlerterms repräsentieren.

### 3.2.3 Fehler-Korrektur Ansätze als alternativer Ansatz zur Berücksichtigung autokorrelierter Störterme

Wenn der ADF-Test die (Trend-)Stationarität eines Indikators bestätigt, ist kein Anlass gegeben, die DAS Monitoring Ergebnisse bezüglich des Scheinkorrelations-Risikos in Frage zu stellen. Nachdem in Abschnitt 3.1.2.2 aber darauf hingewiesen wurde, dass die Ergebnisse von DAS-Trendanschätzungen mit autokorrelierten Residuenstrukturen durch die ergänzende Schätzung von Fehler-Korrektur Modellen substanziell überprüft werden können, spezifiziert der im Tool implementiert Algorithmus für sämtliche I(0)-Indikatoren automatisch ein entsprechendes Fehler-Korrektur Modell. Die Ergebniszusammenfassung der ADF-Testroutinen kennzeichnet daher sämtliche I(0)-Befunde als „ECM“ (Abkürzung für „Error Correction Model“).

Nach Vorgabe der Anwenderinnen und Anwender werden dabei verschiedene Trendhypothesen in einer automatisierten Spezifikationsroutine automatisch auf ihre Signifikanz geprüft. Um eventuelle Autokorrelationseffekte angemessen berücksichtigen zu können, werden dabei ggf. mehrere verzögerte Realisationen der abhängigen Variablen zusätzlich in die Regressorliste aufgenommen. Die getesteten Spezifikationen werden wiederum automatisch auf Signifikanz sämtlicher erklärender Variablen getestet und ggf. unter Ausschluss nicht signifikanter Regressoren neu geschätzt. Die implizierten Ergebnisse der final ausgewählten Testgleichung werden dann als alternative Projektion des zukünftigen Verlaufs des analysierten Indikators in ihrem Niveau abgebildet, d.h. im Tool erfolgt wiederum automatisch eine Projektion von  $\hat{y}_{t+1} = y_t + \widehat{\Delta y}_t$ , wobei  $\hat{y}$  die Niveau-Projektion der Zeitreihe  $y$  bezeichnet.

Wenn, wie in Abschnitt 3.1.2.2 beispielhaft vorgestellt, die Projektionseigenschaften einer entsprechenden Schätzung kaum von den Projektionseigenschaften auf Basis der DAS Monitoring Befunde abweicht, können diese Befunde insgesamt als ausgesprochen robust angesehen werden. Deutliche Abweichungen zwischen den jeweils prognostizierten Indikatorverläufen verdeutlichen die Sensitivität der jeweiligen Schätzergebnisse und sollten als Anlass zur intensivierten inhaltliche Prüfung der jeweiligen Bewertungsergebnisse angesehen werden.

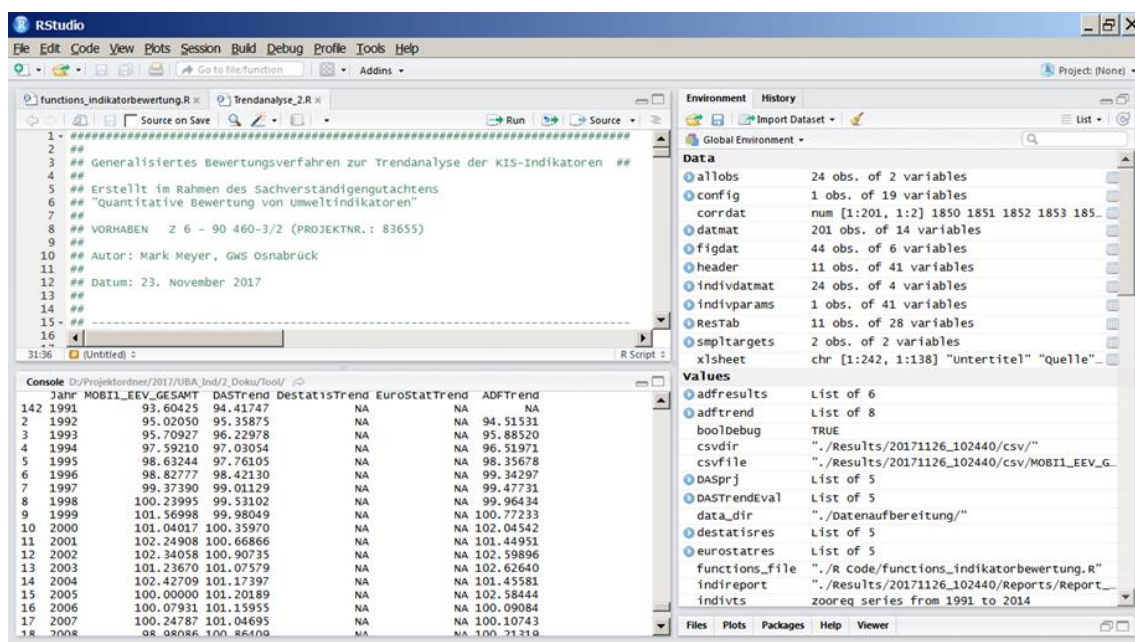
## 4 Abschließende Anmerkungen und Ausblick

Im Rahmen des Sachverständigengutachtens „Quantitative Bewertung von Umweltindikatoren“ wurde ein R-Programm entwickelt, mit welchem alternative Bewertungsverfahren zur Beurteilung der Entwicklungen beliebiger Indikatoren IT-gestützt automatisiert angewandt werden können. Die im Tool implementierten statistischen Analyseroutinen dienen insbesondere zur empirischen Überprüfung folgender Grundthemen der angewandten indikatorbasierten Berichterstattung:

- ▶ Analyse des historischen Verlaufs der Indikatoren (Identifikation von Trendmustern),
- ▶ Fortschreibung zukünftiger Trendentwicklungen,
- ▶ Prüfung der Stabilität identifizierter Trendmuster (Signifikanz geschätzter Schätzfunktionen),
- ▶ Abschätzung zukünftig erwartbarer Zielerreichungsgrade.

Dabei werden den Anwenderinnen und Anwendern des Tools die zuletzt in Schönthaler und von Andrian-Werburg (2015) beschriebenen DAS Monitoring Routinen sowie gängige Bewertungs- und Fortschreibungsmethoden des Statistischen Bundesamtes sowie des Statistischen Amtes der Europäischen Union (Eurostat) zur flexiblen Verwendung zur Verfügung gestellt. Neben diesen etablierten Bewertungsverfahren wurde zudem ein eigenständiger zeitreihenökonomischer Spezifikationsalgorithmus unter automatisierter Berücksichtigung der Ergebnisse von ADF-Tests (Dickey und Fuller 1979) im Tool implementiert. Optional können zudem die Ergebnisse alternativer Trendregressionen betrachtet und die Auswirkungen autonomer Veränderungen der historisch identifizierten Wachstumspfade angezeigt werden. Damit konnte dieses Vorhaben eine wesentliche Grundlage für die zukünftig angestrebte Definition einer standardisierten Methodik der Indikatorbewertung am Umweltbundesamt schaffen.

Abbildung 22: Exemplarische Ansicht einer Benutzerumgebung bei Anwendung des Analysetools



Quelle: Eigene Darstellung auf Basis einer exemplarischen Anwendung der R-Anwendung.

Der Quellcode dieser Anwendung wurde dem Umweltbundesamt zum Projektabschluss vollständig übergeben. Zudem erfolgte eine ergänzende Dokumentation der Funktionalität der Anwendung sowie relevanter technischer Parameter zur Konfiguration alternativer Funktionalitäten in einem eigenständigen Manual. Da diese Anwendung in R programmiert wurde (siehe Abbildung 22 für einen beispielhaften Screen-Shot), kann das Programm auf beliebigen Arbeitsplatz-PCs des Umweltbundesamtes ohne größeren Aufwand zur lizenzfreien Nutzung installiert werden. Damit steht diese Anwendung ab sofort sämtlichen Experten des Umweltbundesamtes zur umfassenden Evaluation zur Verfügung.

Umfassende Feldtests waren im Rahmen dieses Sachverständigengutachtens nicht vorgesehen. Es kann daher erwartet werden, dass bei einer intensiven Nutzung dieses Prototypen die jeweiligen Anwenderinnen und Anwender noch wertvolle Rückmeldungen zur Steigerung der Funktionalität der Anwendung im Hinblick auf individuelle Analysefragen geben können. Im Hinblick auf zukünftig denkbare Weiterentwicklungen des Tools sollte versucht werden, diese systematisch zu erfassen und nachhaltig zu dokumentieren. So wäre es dann bspw. denkbar, in möglichen Folgeprojekten moderierte Workshops vorzubereiten, in denen entsprechend identifizierte Themenschwerpunkte mit den bisherigen Anwenderinnen und Anwendern vertieft diskutiert werden könnten, um einen Zielgruppen-spezifischen Ausbau der aktuellen Funktionalität des derzeitigen Prototypen voranzutreiben.

Neben entsprechenden Folgeaktivitäten zur Steigerung der Handhabbarkeit der Anwendung verweisen die in diesem Sachverständigengutachten aufgeführten methodischen Anmerkungen auf naheliegende Ansatzpunkte für eine inhaltliche Weiterentwicklung dieses Tools. Kurzfristig wird diesbezüglich zunächst eine umfassende Überprüfung der Eigenschaften der exemplarisch umgesetzten eigenen Bewertungsroutinen empfohlen. Da dieses Sachverständigengutachtens nicht für umfassende methodische Grundlagenforschungen intendiert war, konnten diese Bewertungsroutinen bislang insbesondere noch nicht hinsichtlich folgender inhaltlicher Fragestellungen weiter evaluiert werden:

- ▶ Empirische Überprüfung und Verbesserung des derzeitigen Testniveaus: Da die prototypische Umsetzung des Trendidentifikationsalgorithmus für die ersten Differenzen der analysierten Zeitreihen eine mehrstufige Testroutine darstellt, sollte das daraus resultierende empirische Signifikanzniveau in ergänzende Simulationsstudien konkret analysiert werden. Im Rahmen dieses Vorhabens wurde nachgewiesen, dass bei individueller Anwendung eines ADF-Tests auf simulierte Random Walk-Realisationen theoretisches sowie empirisches Signifikanzniveau (im Rahmen üblicher statistische Schwankungen) miteinander übereinstimmen. Damit kann derzeit keine robuste Aussage bezüglich des empirischen Testniveaus der derzeitigen sequentiellen Vorgehensweise getroffen werden, welche zunächst die DAS Monitoring Routinen anwendet, auf Basis dieser Befunde einen ADF-Test durchführt und daraufhin nach Maßgabe der ADF-Ergebnisse eine statistische Trendidentifikation vornimmt. Es ist aber zu erwarten, dass das empirische Testniveau dieser Routine durch eigenständige Weiterentwicklungen noch durchaus gesteigert werden kann.
- ▶ Empirische Überprüfung und Verbesserung der Trennschärfe des derzeitigen Testansatzes: Die Anwendung des ADF-Tests war im Rahmen dieses Vorhabens naheliegend, da (wie im Gutachten gezeigt wird) der ADF-Testansatz unmittelbar aus den in den DAS Monitoring Routinen unterstellten Trendhypothesen hergeleitet werden

kann. In verschiedenen zeitreihenökonometrischen Anwendungen wurde in der Vergangenheit allerdings bereits die relativ geringe Trennschärfe dieses Tests kritisiert. Dies bedeutet, dass der ADF-Test bei Anwendung auf nicht-stochastische Trendreihen häufiger irrtümlich signalisiert, dass diese Zeitreihe stochastische Trendeigenschaften aufweist. Das numerische Ausmaß dieses Effekts könnte für die derzeitige Implementation des ADF-basierten Trendidentifikationsalgorithmus wiederum durch ergänzende Simulationsstudien abgeschätzt werden. Sollte sich dieser Effekt dabei als nicht vernachlässigbar erweisen, sollten zur Verbesserung der derzeitigen Trennschärfe dieses Testalgorithmus Vergleichsstudien unter Verwendung alternativer Einheitswurzeltests durchgeführt werden.

Wichtig erscheint es auch darauf hinzuweisen, dass dynamische Trendspezifikationen, also Fehler-Korrektur Modelle des eigenen Ansatzes wie auch DAS Monitoring Befunde mit autokorrelierten Störtermen, derzeit im Tool statisch fortgeschrieben werden. Dies bedeutet, dass in diesen Fällen zur Abbildung der Trendverläufe unterstellt wird, dass die zukünftigen Störterme stets einen Wert von Null annehmen. Folglich ist es derzeit nicht möglich, die inhärente Sensitivität der Trendprojektionen bezüglich der zukünftigen Entwicklung der Störterme abzubilden. Bei zukünftigen Überarbeitungen des Tools bieten sich damit insbesondere folgende Erweiterungen an:<sup>38</sup>

- ▶ Einbau eines dynamischen Projektionsmechanismus: Gemäß der geschätzten Verteilungsparameter des Störterms können für jede Projektion auch zufällige Störtermausprägungen generiert werden. Unter Einbindung eines solchen Mechanismus wäre es dann auch möglich, eine Funktion zur automatischen Generierung zahlreicher dynamischer Projektionen zu generieren, um auf Basis der so beobachtbaren Verteilung der dynamischen Projektionswerte die Bandbreite der zukünftig erwarteten Indikatorentwicklungen aufzeigen zu können.
- ▶ Erweiterte Abschätzung der Zielerreichungsgarde: Wenn ein entsprechender dynamischer Projektionsmechanismus im Tool implementiert wird, kann aufgrund der simulierten Bandbreite der zukünftig erwarteten Indikatorentwicklungen auch die bisherige Abbildung von Zielerreichungsgraden aussagefähig erweitert werden. Während das Tool in den hier betrachteten Fällen derzeit lediglich Aussagen darüber treffen kann, ob ein vorgegebener Zielwert (unter Ausblendung zukünftiger zufälliger Störeffekte) wahrscheinlich erreicht wird, kann dann (unter Berücksichtigung der zukünftig zu erwartenden Störeffekte) die Wahrscheinlichkeit der zukünftigen Zielerreichung anhand der simulierten Indikatorverläufe konkret quantifiziert werden.

Daneben verweisen die in diesem Sachverständigengutachten aufgeführten methodischen Anmerkungen auf weiterführende Aspekte, welche eine Projektion historischer Indikatorbeobachtungen nachhaltig beeinflussen können. Um mittelfristig eine standardisierte Methodik der Indikatorbewertung etablieren zu können, wird insbesondere eine weitergehende

---

<sup>38</sup> Auf Basis ähnlicher Überlegungen sollten natürlich auch für statische Projektionen die Möglichkeiten zur Abbildung von Konfidenzintervallen geprüft werden.



methodische Abwägung folgender Herausforderungen in zukünftigen Forschungsprojekten empfohlen:

- ▶ Systematische Abwägung der Handlungsalternativen zur Identifikation von Strukturbrüchen: Nachdem das Gutachten darauf hinweist, dass Strukturbrüche den Verlauf projizierter Trendfortschreibungen deutlich beeinflussen können, repräsentiert die Einbindung expliziter Strukturbruchtests eine naheliegende Option zur zukünftigen Weiterentwicklung der Funktionalität des Tools. Tatsächlich finden sich in der akademischen Literatur zahlreiche Vorschläge zur datenbasierten Identifikation von Strukturbrücheffekten. Angesichts des üblicherweise relativ geringen Beobachtungsumfangs der KIS-Indikatoren kann eine verlässliche Abwägung der tatsächlichen Eignung dieser Verfahren zur Anwendung im Tool allerdings nur anhand entsprechender zusätzlicher Simulationsstudien abgeschätzt werden.
  
- ▶ Robuste Fortschreibung nichtlinearer Trenddynamiken: Dieses Gutachten rät von einer Projektion höhergradiger Trendpolynome zur zukünftigen Fortschreibung nichtlinearer Trendentwicklungen ab. Während im gegebenen Projektrahmen auf eine explizite Vorstellung alternativer Ansätze zur expliziten Modellierung nichtlinearer Zeitreihendynamiken verzichtet wurde, sei zunächst darauf hingewiesen, dass entsprechende Methoden durchaus verfügbar sind. Vor Hintergrund eigener Erfahrungen mit der Identifikation und Schätzung von nichtlinearen Zeitreihenmodellen (siehe hierzu bspw. Maringer und Meyer 2008), scheinen diese grundsätzlich zur automatisierten Identifikation von Strukturbrücheffekten wie auch zur Prognose von zukünftigen Indikatorentwicklungen geeignet zu sein. Es erscheint daher empfehlenswert, zumindest für ausgewählte Indikatoren im Rahmen eines eigenständigen Forschungsprojektes die konkreten Prognoseeigenschaften einer entsprechenden Methode vertieft zu diskutieren.

Während dieses Vorhaben der erstmaligen konkreten Realisation der Anwendung diene, verbleibt somit umfassender Raum für zukünftige weitergehende Entwicklungsaktivitäten. Da es sicherlich wünschenswert wäre, sowohl die Grundlagenforschung zu den implementierten Methoden als auch den Zielgruppen-orientierten Ausbau des Tools in einem harmonisierten Vorhaben gemeinsam weiterzuverfolgen, sollte hierzu allerdings eine deutlich längere Bearbeitungszeit erwogen werden.

Bei Abwägung der Potenziale entsprechender Folgeaktivitäten sollte abschließend auch darauf hingewiesen werden, dass die derzeitige Realisation in R im gegebenen Projektrahmen naheliegend war, dies aber nicht als grundsätzliche Festlegung der Softwarebasis für weitergehende Toolentwicklungen verstanden werden sollte. So war diese Entscheidung maßgeblich dadurch beeinflusst, dass die Priorität im Vorhaben auf der Realisierung maximaler Funktionalität lag. Der zeitliche Entwicklungsaufwand zur Umsetzung der für dieses Sachverständigengutachten benötigten Funktionalität wurde daher nicht wesentlich durch eine Entscheidung zwischen R und alternativen (aber in der Regel lizenzpflichtigen) zeitreihenökonometrischen Softwarepaketen beeinflusst. Diese Entscheidungssituation kann sich allerdings bei eventuellen späteren Folgeprojekten vollkommen anders darstellen, da in professionellen Anwendungsumgebungen häufig bereits viele Funktionalitäten angelegt sind,

welche bei einer Umsetzung in R zunächst (zumindest in einem nennenswerten Umfang) eigenständig programmiert werden müssen.

So verfügt bspw. das zeitreihenökonometrische Softwarepaket EViews (im Gegensatz zu R) bereits über eine interaktive grafische Benutzeroberfläche. Bei der Entwicklung eigener Anwendungen können daher in EViews Anwendungs-spezifische Auswahlmenüs sowie weitergehende Konfigurationsdetails relativ flexibel als interaktive Dialogelemente angelegt werden. Diverse Indikatoransichten (bspw. in tabellarischer Form, als Zeitreihenplot, Histogramm, ...) können standardmäßig per Knopfdruck generiert sowie in vielfältigen Dateiformaten exportiert werden (bspw. als \*.eps, \*.jpg, \*.png oder \*.pdf - Dateien). Daher ist davon auszugehen, dass bspw. der notwendige Arbeitsaufwand zur Konzeption und Programmierung einer automatisierten grafischen Benutzerführung, welche die Funktionalität der in diesem Vorhaben erstellten R-Anwendung in interaktiver Form zur Verfügung stellt, in EViews mit einem Arbeitsvolumen weniger Personenmonate realisiert werden könnte.

## 5 Quellenverzeichnis

- Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6): 716–723.
- Blumers, M. und Kaumanns, S. (2017): Neuauflage der deutschen Nachhaltigkeitsstrategie. *Wirtschaft und Statistik* 2017(1), 96-109.
- BMUB – Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (2015): Indikatorenbericht 2014 zur Nationalen Strategie zur biologischen Vielfalt; Download unter <http://www.bmub.bund.de/service/publikationen/downloads/details/artikel/indikatorenbericht-2014/>.
- Cochrane, D. und Orcutt, G. (1949): Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44, S. 32–61.
- Dickey, D. und Fuller, W. (1979): Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, S. 427–431.
- Europäische Union (2015): Sustainable development in the European Union. 2015 monitoring report of the EU sustainable development strategy.
- Europäische Union (2017): Sustainable development in the European Union. Monitoring report on progress towards the SDGs in an EU context. Download unter: <http://ec.europa.eu/eurostat/en/web/products-statistical-books/-/KS-04-17-780>.
- Eurostat (2014): Getting messages across using indicators – A handbook based on experiences from Sustainable Development Indicators. Luxembourg
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. und Tutz, G. (2016): *Statistik. Der Weg zur Datenanalyse*. 8. Auflage, Berlin Heidelberg, Springer.
- Granger, C. und Newbold, P. (1974): Spurious regressions in econometrics. *Journal of Econometrics*, 2, 111–120.
- Langsdorf, S. und Hirschnitz-Garbers, M. (2014): Looking to the future: Trend report for a forward-looking resource policy. Download unter: <https://www.umweltbundesamt.de/publikationen/looking-to-the-future>.
- Martone, O. (2013): Die Methode der Trendanalyse. In: Arbeitskreis Umweltökonomische Gesamtrechnungen der Länder (2013): *Umweltökonomische Gesamtrechnungen der Länder*. Band 3. Analysen und Berichte. Methodische Weiterentwicklungen in den UGRdL. Ausgabe 2013. Download unter: [http://www.ugrdl.de/pdf/ugrdl\\_analyse\\_2013.pdf](http://www.ugrdl.de/pdf/ugrdl_analyse_2013.pdf)
- MacKinnon, J. (1996): Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11, S. 601-618.
- Maringer, D. und Meyer, M. (2008): Smooth Transition Autoregressive Models -- New Approaches to the Model Selection Problem. *Studies in Nonlinear Dynamics & Econometrics*, 12(1), doi:10.2202/1558-3708.1469.
- Popper, K. (1994): *Logik der Forschung*. Tübingen: Mohr & Siebeck, 10. Auflage.
- Schönthaler, K. und von Andrian-Werburg, S. (2015): *Handbuch zur Verstetigung der indikatorbasierten Berichterstattung zur Umsetzung der Deutschen Anpassungsstrategie an den Klimawandel (DAS)*. Stand: 26.01.2015. Download unter: <https://www.umweltbundesamt.de/dokument/handbuch-zur-verstetigung-der-indikatorbasierten>.
- Statistisches Bundesamt (2017): Nachhaltige Entwicklung in Deutschland, Indikatorenbericht 2016. Download unter: [https://destatis.de/DE/Publikationen/Thematisch/UmweltökonomischeGesamtrechnungen/Umweltindikatoren/IndikatorenPDF\\_0230001.pdf?\\_\\_blob=publicationFile](https://destatis.de/DE/Publikationen/Thematisch/UmweltökonomischeGesamtrechnungen/Umweltindikatoren/IndikatorenPDF_0230001.pdf?__blob=publicationFile).
- UBA – Umweltbundesamt (2015): Monitoringbericht 2015 zur Deutschen Anpassungsstrategie an den Klimawandel. Download unter: <https://www.umweltbundesamt.de/publikationen/monitoringbericht-2015>.



UBA – Umweltbundesamt (2017): Daten zur Umwelt 2017. Download unter:  
<https://www.umweltbundesamt.de/publikationen/daten-zur-umwelt-2017>.

Winker, P. (2010): Empirische Wirtschaftsforschung. Berlin Heidelberg, Springer