ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Rendtel, Ulrich

Working Paper

Die Zukunft der Statistik: Eine persönliche Betrachtung

RatSWD Working Paper, No. 166

Provided in Cooperation with:
German Data Forum (RatSWD)

Suggested Citation: Rendtel, Ulrich (2010) : Die Zukunft der Statistik: Eine persönliche Betrachtung, RatSWD Working Paper, No. 166, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:
https://hdl.handle.net/10419/75310

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Rat für Sozial- und
WirtschaftsDaten (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper Nr. 166

Die Zukunft der Statistik: Eine persönliche Betrachtung

Ulrich Rendtel

Dezember 2010

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Kontakt: Rat für Sozial- und Wirtschaftsdaten | Mohrenstraße 58 | 10117 Berlin | office@ratswd.de

Die Zukunft der Statistik: Eine persönliche Betrachtung[1]

Ulrich Rendtel

Freie Universität Berlin (Ulrich.Rendtel@fu-berlin.de)

Zusammenfassung:

Anhand eines persönlichen Rückblicks auf die Entwicklungen in den Bereichen Rechnerentwicklung, Datenzugang und Entwicklung von Statistik-Software werden Trends für die zukünftige Entwicklung der Statistik im Bereich der Wirtschafts- und Sozialwissenschaft hergeleitet. Insbesondere werden die Rolle von R, neue Möglichkeiten des Datenzugangs, das Verhältnis zur Amtlichen Statistik und die Einführung neuer Studiengänge im Bereich der Statistik angesprochen. Die Darstellung bezieht sich auf den Bereich der Wirtschafts- und Sozialwissenschaften. In anderen Wissenschaftsbereichen, wo die Statistik als Biometrie, Psychometrie etc. firmiert, mögen die hier dargestellten Entwicklungstendenzen irrelevant sein.

Schlüsselwörter: Datenzugang, Statistik-Pakete, R, Amtliche Statistik, Statistik-Studiengänge

JEL-Klassifikation: C10

1 Ich danke Gert Wagner und Jürgen Wolters sowie allen Teilnehmern des Festcolloquiums „Die Zukunft der Statistik“ (22.10.2010 im Clubhaus der Freien Universität Berlin) für hilfreiche Kommentare und gelegentlichen Widerspruch.

1. Einleitung

Die Zunft der Statistiker[2] wird in der öffentlichen Meinung zwiespältig angesehen. Im negativen Fall rangieren sie als Erbsenzähler oder gar Lügner. Bei positiver Bewertung sind es die Statistiker, die die Basis für eine erfolgreiche Politikberatung schaffen. Sie beherrschen nämlich die Kunst, aus abstrakten Zahlenkolonnen konkrete Informationen zu extrahieren. Gerne werden Statistiker in die zweite Reihe der Wissenschaften als Vertreter einer sogenannten Hilfswissenschaft gestellt, wenn sie nicht ganz vergessen werden. An den Sozial- und Wirtschaftswissenschaftlichen Fakultäten führen die Substanzwissenschaftler, also die Ökonomen oder die Soziologen das Wort. Das merkt man spätestens an den Rankings der Zeitschriften bei der Vergabe der universitären Leistungsmittel, wo JASA[3] meist nicht als A-Journal auftaucht und der Professur entsprechend weniger Geld einbringt. Weitere der Statistik zugeschriebene Hilfsdienste sind von fragwürdiger Qualität. Da firmieren die Statistik-Pflichtveranstaltungen des Grundstudiums als Rausschmeißer für „unfähige“ Studenten, die nicht rechtzeitig erkannt haben, wie wichtig angeblich die „Beherrschung von Formeln“ für das weitere Studium ist, vgl. etwa Wagner/Büning (2008).

Einige Kollegen interpretieren Veränderungen und Trends im Umfeld der Statistik als Niedergang eines Fachs, z.B. v.d.Lippe/Kladobra (2008). Eckpunkt dieser negativen Sicht ist eine wahrgenommene Abnahme der Vertreter des Fachs Statistik unter den Hochschullehrern und eine Schwächung des verbindlichen Curriculums beim Übergang auf Bachelor- und Masterstudiengänge. (Rendtel 2008, 2009). Andere Fachvertreter tendieren in das gegenteilige Extrem und attestieren sich öffentlich selbst ein dickes Lob („Wir sind gut!“ konnte man beispielsweise im Berliner TAGESSPIEGEL lesen).

Ziel der Ausführungen ist hier jedoch nicht ein Abgesang auf den Untergang einer hohen Methodenkultur. Vielmehr sollen Aussagen auf zukünftige Trends des Fachs Statistik hergeleitet werden. Für das Verhältnis zur Amtlichen Statistik werden sogar konkrete Handlungsempfehlungen gegeben.

---

2 Die Darstellung bezieht sich auf den Bereich der Wirtschafts- und Sozialwissenschaften. In anderen Wissenschaftsbereichen, wo die Statistik als Biometrie, Psychometrie etc. firmiert, mögen die hier dargestellten Entwicklungstendenzen irrelevant sein. Beispielsweise wird man im Bereich der Biometrie anders als in der Wirtschaftswissenschaft immer mit geringen Fallzahlen konfrontiert sein.

3 JASA steht für Journal of the American Statistical Association, sicher eines der Top-5 Journale der Statistik.

Der hier gewählte Ansatz ist zunächst retrospektiv und analysiert aus der Sicht des Autors, was das Fach Statistik in den letzten drei Dekaden bewegt hat. Aus diesen Entwicklungslinien soll extrapoliert werden, welche Trends für die Zukunft des Fachs Statistik zu erwarten sind.

2. Eine Retrospektive der Statistik

2.1 Rechnerentwicklung

Die Entwicklung der letzten 30 Jahre ist gekennzeichnet durch die Verfügbarkeit von immer leistungsfähigeren und preiswerteren Rechnern.[4] Während noch in den 1970er-Jahren beispielsweise SAS nur auf wenigen zentralen Grossrechnern verfügbar war, sind heute alle gängigen Statistikpakete auf jedem Laptop oder PC in kürzester Zeit installierbar. Und auch der Installationsaufwand hat sich stark reduziert. Während man in den 1980er-Jahren noch ca. 50 Installationsdisketten einlesen musste, um SAS zu installieren, benötigt man heute eine bis zwei Installations-CDs mit komfortabler Programmführung.

Diese preiswerte Rechnerkapazität beförderte die Nachentwicklung von Software für viele nicht-lineare Schätzer, die die mathematische Statistik schon in den 30er-Jahren des letzten Jahrhunderts entwickelt hatte, so z.B. Fisher und seine allgemeine Theorie der Maximum Likelihood (ML)-Schätzer, vgl. Fisher (1925). Mittlerweile hat die Verfügbarkeit von hoher Rechnerkapazität neue Methodenansätze hervorgebracht. Es ist sicher kein Zufall, dass Efron den rechnerintensiven Bootstrap in den 1980er-Jahren entwickelte. Generell hat die Nicht- oder Semiparametrische Statistik im Laufe der letzten Jahre an Bedeutung gewonnen. Es hat den Anschein, dass sich die Gewichte zwischen „Brain Power“ und „Computer Power“ deutlich zugunsten des Computers verschoben haben und die mathematische Statistik als Exponentin der „Brain“-Fraktion an Bedeutung verloren hat. Diese etwas grobe Sichtweise muss jedoch relativiert werden. Es stimmt zwar, dass Aussagen über gleichmäßig beste Tests, Effizienzabschätzungen oder das Neyman-Pearson Lemma außer Mode gekommen sind.

4 Der Zeitraum ab 1980 markiert die Periode, seit der der Autor die akademisch verfasste Statistik aktiv mitverfolgt hat. Die Periode vor 1980 war von noch radikaleren Technikumwälzungen geprägt. Vor 1980 gab es noch Lochstreifen- und Lochkartenverarbeitung in Hochschulrechenzentren.

Gleichwohl ist die mathematische Behandlung des intuitiv einleuchtenden Konzepts des Bootstrap alles andere als einfach. [5]

2.2 Datenzugang

Ein weiteres wichtiges Moment ist die Verfügbarkeit von Speichermedien für die Datenhaltung. Diese Entwicklung scheint bei weitem noch nicht abgeschlossen: Kilo-, Mega-, Giga- und jetzt Terabyte! Diese Entwicklung ermöglichte die Haltung von Datensätzen auf lokalen Rechnern und somit die Distribution von sogenannten Scientific Use Files. Noch in den 1990er-Jahren war es notwendig, die Daten der Längsschnittstudie des sozio-oekonomischen Panels (SOEP) aus Platzspargründen als ein komprimiertes Datenfile[6] zu halten und auf diesem gepackten Datensatz Auswertungen durchzuführen. Heute ist diese platzsparende, jedoch einengende Auswertungstechnik überflüssig.

Diese unvorstellbar großen Kapazitäten in der Datenhaltung hat in der Wissenschaft einen regelrechten Datenhunger erzeugt. In einem zehnjährigen Prozess entstand im Bereich der Sozial- und Wirtschaftswissenschaften eine ganz neue Infrastruktur von Forschungsdatenzentren (siehe www.ratswd.de/dat/fdz.php). Aufgabe dieser Forschungsdatenzentren, die ursprünglich im Umkreis der Amtlichen Statistik sowie der Bundesagentur für Arbeit und der Deutschen Rentenversicherung entstanden, ist die Umsetzung der Datenschutzgesetzgebung, die der Wissenschaft – im Gegensatz zu nicht-wissenschaftlichen Einrichtungen – einen Zugang zu sogenannten faktisch anonymisierten Mikrodaten einräumt. Seit dem Jahr 2004 existiert der Rat für Sozial- und Wirtschaftsdaten (RatSWD), dessen Anliegen die Verbesserung der „informationellen Infrastruktur“ ist (siehe www.ratswd.de). Unter diesem abstrakten Begriff verbirgt sich nicht nur ein verbesserter Datenzugang. Er deckt auch die Ausbildung zum Umgang mit diesen Daten ab und wird mit zahlreichen Workshops an Universitäten und Forschungsdatenzentren verbunden. Die Einrichtung einer Vermittlungsstelle zwischen Datenproduzenten und Datenauswertern auf nationaler Ebene existiert in Europa nur noch im Mutterland der Statistik, dem Vereinigten Königreich (UK). Dort hat sich der Economic and Social Research Council (ESRC) zu einer wichtigen Schaltstelle für die Statistik als Instrument der Datenanalyse entwickelt.

---

5 Unter Konvergenz in Verteilung geht es nicht, vgl. z.B. Davison/Hinkley (2005).
6 Das Komprimierungsprogramm RZ00 erreichte je nach Daten eine Platzersparnis von 80 %.

Mittlerweile wurden vom RatSWD 17 Service- und Forschungsdatenzentren akkreditiert (Stand: Dezember 2010). Neben der Distribution von Scientific Use Files werden auch Campus-Files für Ausbildungszwecke (Zwick 2008) sowie Plätze für Gastwissenschaftler angeboten. Die Zahl von gemeinsamen Workshops der neuen Forschungsdatenzentren und der Universitäten wächst stetig, vgl. z.B. „SOEP at Campus“ (www.unibielefeld.de/soz/soep@campus).

Insgesamt hat die Datenorientierung des Fachs Statistik enorm zugenommen und ich prognostiziere eine Fortdauer dieses Trends.

3. Substantielle Fragestellungen als Motor der Statistik

Es liegt tief im Wesen der Statistik, eine angewandte Wissenschaft zu sein, d.h. einzelne Problemstellungen aus diversen wissenschaftlichen Disziplinen werden mit formalen statistischen Methoden behandelt und führten auch zu einer Weiterentwicklung des Methodenapparats der Statistik. Welche Problemstellungen dies sind, ist schwer vorherzusagen. Allerdings scheint eine hohe Verfügbarkeit über Geld und Ressourcen eine wesentliche Voraussetzung für die Etablierung entsprechender Forschungsaktivitäten zu sein. Ersatzweise helfen auch Forschungsaufträge durch die Legislative. Hierzu einige Beispiele:

- Der finanzstarke Bankbereich eröffnete mit dem Basel-II-Abkommen und der Fixierung des Value-at-Risk die Suche nach nicht normalverteilten abhängigen Zufallsgrößen. Dies führte zur Wiederentdeckung der Copulae, Schmid/Trede (2005). Immerhin ist jetzt der Begriff des Quantils im Bankbereich verankert. Der zweite Schub des heute kurz als „Finance“ titulierten Bereichs kommt aus der Welt der Finanzderivate und der Bewertung dieser Kurswetten. Dies führte zu einem deutlichen Aufschwung in der Modellwelt der stochastischen Prozesse. Derzeit sind diese Fragestellungen so „in“, dass selbst eingefleischte BWL-Studenten vor dem Ito-Kalkül stochastischer Differentialgleichungen nicht zurückschrecken. Allerdings ziehen es die Banken vor, für die Lösung der aus diesem Kalkül abgeleiteten partiellen Differentialgleichungen Physiker einzustel-

len. Mittlerweile hat sich der vor 30 Jahren noch völlig unbekannte Begriff „Financial Statistics“ fest auf Konferenzen und bei Lehrstühlen etabliert.

– Eine Statistik-Förderung mit klarem politischem Auftrag war die Hartz-Gesetzgebung zur Arbeitsplatzförderung. Hier gab es den gesetzlichen Auftrag, die diversen Förderinstrumente wissenschaftlich zu evaluieren.

Dieser Auftrag zur wissenschaftlichen Evaluation von Maßnahmen der Hartz-Gesetzgebung führte zu einem Boom der Analysen von Pseudo-Experimenten durch Matching-Verfahren beziehungsweise Missing Value-Verfahren, vgl. Lechner/Pfeiffer (2007). Der Missing Value-Ansatz beruht auf Rubins Counterfactual Approach, wonach sich der Erfolg einer Maßnahme als Outputdifferenz zweier Werte ergibt, nämlich des Output unter der Maßnahme und des Outputs ohne Maßnahme. Da meist nur ein Output beobachtet wird, liegt ein substantielles Missing Data-Problem vor. Missing Data-Probleme treten wegen der schon erwähnten höheren Verfügbarkeit von Umfragedaten verstärkt als Statistik-Problem auf. Neben den klassischen Gewichtungs- und Kalibrierungsverfahren der Umfrageforschung ist seit den 1980er-Jahren von Rubin die Strategie der multiplen Imputation entwickelt worden, vgl. Rubin (1996). Hier kommt die Bayes’sche Statistik voll zum Einsatz. Wegen der gestiegenen Rechnerleistung scheint auch ein notorisches Problem des Bayes-Kalküls, nämlich die Evaluation von Mehrfachintegralen bei der Berechnung der Posterior-Verteilung kein ernsthaftes Problem mehr. Das Zauberwort heißt hier MCMC (= Markov Chain Monte Carlo) und meint den Einschwingvorgang auf eine gemeinsame Verteilung bei bekannten Konditionalverteilungen, vgl. Gelman/Meng (2004).

Insbesondere lehrt dieses Beispiel den weiten, verschlungenen Pfad von einem außerstatistisch gesetzten Anreiz (Evaluation der Hartz-Reformen), der formalen Einbettung und den oft überraschenden Ähnlichkeiten mit anderen Problemen auf der formalen Ebene (Missing Data-Problem) und der innerstatistischen Behandlung des Problems (Realisation des Erwartungswertes einer Posterior-Verteilung).

Es hat nicht an Versuchen gefehlt, die Behandlung fehlender Daten, sei es bei Umfragen oder bei der Programm-Evaluation, zu automatisieren. Erwähnt seien etwa Raghunathans Programm IVEware (Raghunathan et al. 2002) oder v. Buurens MICE (v. Buuren/Groothuis-Oudshoorn, 2010). Aber wie bei allen harten statistischen Problemen erweisen sich automatische Routinen als wenig

erfolgreich. Automatische Modellsuchverfahren findet man zwar in vielen Bereichen der Statistik (ARMA-Modelle bei Zeitreihen, Loglineare Modelle bei Kontingenztabellen). Trotzdem werden diese Strategien in der Praxis nur selten angewendet, weil man der Zufälligkeit des gewählten Modells misstraut. Dieses Schicksal würde ich auch für die Automationsversuche im Bereich von Missing Data vorhersagen.

4. Erfolgreiche und nicht erfolgreiche Ansätze

Während der letzten drei Dekaden hat die Statistik viele Ansätze hervor gebracht. Eine Erfolgsgeschichte war die Einführung und vor allem die softwaremäßige Umsetzung von verallgemeinerten linearen Modellen, die mittlerweile zum Standardstoff jeder weiterführenden Statistikausbildung gehören. Wohl etabliert sind mittlerweile auch diverse Hazardraten-Modelle zur Analyse von Lebensdauern. Diese entstammten ursprünglich der Biometrie sowie der Analyse der Lebensdauer technischer Systeme. Aber seit der Verfügbarkeit von individuellen Daten über die Dauer von Arbeitslosigkeit sind sie auch zum festen Bestandteil an wirtschaftswissenschaftlichen Fakultäten geworden.

Es gab aber auch Ansätze, die sich nach meinem Eindruck nicht durchgesetzt haben. Gefährdet sind hier in erster Linie Ansätze, die die Arbeit des Statistikers wegautomatisieren wollten. Wer kennt noch die sogenannten „Expertensysteme“, die den Arzt bei der Diagnose unterstützen sollten? Formal handelte es sich um ein System von Bayes-Vorinformationen, das durch Beobachtungen fortgeschrieben wurde. Es gab viele Anläufe, die aber meist enttäuschend endeten.

Der Ansatz automatischer Modellsuchstrategien wurde zwar auf den gängigen Softwarepaketen realisiert. Aber die Anzahl der Kollegen, die ihre Modelle auf diese Art suchen, ist sehr begrenzt. Man misstraut der Zufälligkeit dieser Auswahlstrategien.

Irgendwie sind diese Resultate beruhigend, zeigen sie doch, dass die Tätigkeit eines Statistikers kaum wegzurationalisieren ist. Meine persönliche Definition von Statistik lautet „Daten und Denken“. Der zweite Teil dieses Tätigkeitsfeldes lässt sich schwerlich automatisieren.

5. Die Beziehungen der Statistik zu Nachbardisziplinen

Die Nachbarn heißen Pattern Recognition, Data Mining, Neuronale Netze oder gar Statistical Computing. Letzteren Begriff würde man naiverweise als Statistik übersetzen. Es geht jedoch um etwas anderes: Es ist die Kunst, aus großen Datenbanken unter Nebenbedingungen einfache beschreibende Maße in Echtzeit herauszufinden. Die Wirtschaftsinformatiker verwenden dafür das Akronym OLAP (Online Analytical Processing).

Pattern Recognition ist eine klassische Domäne der Marktforschung. Hier sind in den letzten Jahren neue Datentypen entstanden, die in völlig neuen Größenordnungen anfallen. Die Scannerkassen eröffneten das Feld der Warenkorbanalyse. Das Internet liefert neue Datentypen, die einerseits Charakteristika der Verbindungen als auch Textelemente enthalten. Zudem ist die Fallzahl in der Regel sehr hoch. Dies hat zu dem neuen Begriff des „Text Mining“ geführt, der sich an das schon gebräuchliche „Data Mining“ anlehnt. Das Data Mining ist eine Sammlung von verschiedenen Strategien zur Mustererkennung und zur Prognose. Der Anwendungsbereich liegt häufig im Bereich der BWL, weshalb der Data Miner von SAS auch „Enterprise Miner“ heißt, vgl. Matignon (2007).

Bei der Prognose wurde ein alter Traum aus dem Bereich der künstlichen Intelligenz verwirklicht: Die Nachbildung der Tätigkeit des menschlichen Gehirns auf dem Rechner. Hier wird eine dezidiert nicht-statistische Sprechweise benutzt, die mitunter Statistiker in Verwirrung stürzt, weil ein Bias in dieser Welt die Konstante eines Neurons ist. Eine exzellente Rückübersetzung in die Welt der Statistik bieten Hastie/Tibshirani (2001) an.

Es brauchte einige Zeit, bis die Statistiker merkten, dass man Neuronale Netze als nicht-lineare Regressionsmodelle interpretieren kann und diese der Gefahr des Overfitting ausgesetzt sind, vgl. Cheng/Titterington (1994). In umgekehrter Richtung wird die Fähigkeit der Statistik zur Prognose jedoch seltener wahrgenommen. Es ist ja nicht das erste Mal, dass das Rad neu erfunden wurde. Was einen Statistiker dabei stutzig macht, ist die geringe Macht der Statistik im Wissenschaftsbetrieb, die nicht einmal zur Informatik herüber reicht.

In diesem Umfeld stellt sich auch die Frage: Wo fängt Statistik an und wo hört sie auf? Klassifikations- und Regressionsbäume (CART) scheinen klar auf ein statistisches Verfahren hinzuweisen, in diesem Fall auf die Regression. Allerdings

folgen die Konstruktionsalgorithmen dieser Bäume ausschließlich Heuristiken und nicht etwa statistischen Ansätzen wie Maximum-Likelihood oder Kleinste Quadrate. Der Preis für die Negierung des statistischen Rahmens ist der Mangel an statistischen Schlussweisen: Testen, Konfidenzintervalle, Konsistenzaussagen – all dies fehlt und kann nur durch die Heuristik aufgewogen werden. Dieser Mangel scheint in unserer wissenschaftlichen Nachbarschaft kaum wahrgenommen zu werden. Man scheint nichts zu vermissen. Es ist ein Befund, der einen nachdenklich werden lässt.

Die Statistiker sollten sich ernsthaft fragen, warum derart souverän auf statistische Essentials verzichtet wird: Ganz offensichtlich, weil die Nutzer über ihre Techniken substantiell gut interpretierbare Ergebnisse gewinnen.

Was lernen wir daraus? Es gilt immer noch das Verdikt über statistischen Modelle: „Alle Modelle sind falsch. Aber einige Modelle sind nützlicher als andere!“ Der Begriff der „Nützlichkeit“ ist **kein** statistischer Begriff. Er orientiert sich an der Interpretation der Ergebnisse, die durch den Anwender und nicht durch den Statistiker definiert wird. Eine derartige Orientierung auf den Datennutzer fällt nicht vom Himmel. Sie muss erlernt werden. Allerdings ist diese Nutzerorientierung nur in seltenen Fällen Gegenstand der Ausbildung von Statistikern. Eine notwendige Voraussetzung hierzu ist die Existenz einer universitären Statistischen Beratungseinheit[7], vgl. z.B. das STABLAB an der Ludwigs-Maximilian-Universität München (www.stat.uni-muenchen.de/stablab) oder fu:stat an der Freien Universität Berlin (www.stat.fu-berlin.de).

Weiterhin ist Arroganz gegenüber dem vermeintlich unstatistischen Vorgehen unserer Kollegen unangebracht, weil auch im eigenen Haus gar nicht so selten gegen statistische Regeln verstoßen wird, etwa beim Picken nach Signifikanzen[8] von einzelnen Merkmalen bei der Modellsuche.

7 In früheren Zeiten hatten die Universitätsrechenzentren häufig auch die Aufgabe, Nutzer nicht nur bei Problemen mit Software zu beraten, sondern ihre Nutzer auch bei der Analyse von Statistischen Daten zu beraten. Allerdings war das Servicespektrum dieser Rechenzentren viel zu groß, um solche Aufgaben effizient lösen zu können, und nicht selten fehlte es den Beratern an einer hinreichenden statistischen Expertise.

8 Gemeint ist das Durchsuchen von Listen mit geschätzen Regressionskoeffizienten nach Signifikanzen.

6. Die Entwicklung von Statistik-Software

Rückblickend ist die Entwicklung der Auswertungssoftware, der sogenannten „Statistik-Pakete“, geradezu rasant gewesen. Noch in den 1970er-Jahren wurde mit Lochkarten gearbeitet. In den 1980er-Jahren wuchsen auf den Großrechnern[9] größere Auswertungsprogramme, eben die „Statistik-Pakete“, heran. Sie genossen bald den Ruf, „Dinosaurier“ zu sein: Zu groß, um effizient zu sein – zum Aussterben verurteilt. Allein der Stapel aller zugehörigen Handbücher eines Pakets erreichte beängstigende Ausmaße. Dann kam Ende der 1980er-Jahre GAUSS mit einer voll ausgebauten Matrixsprache auf einem PC heraus. Dies schuf viel Flexibilität und animierte zur Entwicklung eigener Programme, die oft frei unter Wissenschaftlern ausgetauscht wurden. Von hier war es nicht mehr weit bis zum R-Projekt: Eine lizenzfreie Matrixsprache und eine explizite Open Source-Politik bewirkten eine weitere Beschleunigung der Entwicklung von Statistik-Software und ihrer allgemeinen Verfügbarkeit. R ist auf dem Wege, in der akademischen Welt die Nummer eins zu werden und die Standbeine von R in der nicht-akademischen Welt werden immer wahrnehmbarer.

Allerdings sind die sogenannten Dinosaurier der Statistik-Software nicht ausgestorben. SAS und SPSS leben immer noch. Die meterdicken Handbücher sind intelligent gemachten Online-Dokumentationen gewichen, die streckenweise mit gut gemachten Statistik-Lehrbüchern mithalten können. Beispielsweise ist die SAS-Online Dokumentation frei verfügbar (http://support.sas.com/documentation/onlinedoc/91pdf/ index.html). Die Stärke der großen Pakete ist die systematische Aufarbeitung der wichtigsten statistischen Verfahren. Hierin sind sie komplementär zu R (www.r-project.org). Und sie sollten auch für größere Datensätze stabil laufen.[10] Der R-Core hat hingegen immer noch Schwierigkeiten mit größeren Datensätzen und die Sammlung der Open-Source-Programme erfolgt lexikographisch. Für die Dokumentation ist jeder Autor selbst verantwortlich. Der Programmcode liegt zwar für jeden Benutzer offen, aber das Lesen einer Matrixsprache stellt eine gewisse Methodenbarriere dar, die von vielen nicht überwun-

9 Sie waren wirklich groß!
10 Allerdings erlebt man hierbei manchmal böse Überraschungen.

den wird. Dafür ist R umsonst und superschnell in der Bereitstellung von Modellschätzern.

Mittlerweile wurden einige Programme so umgeschrieben, dass R für Anwender mit dem gewohnten Komfort der Statistikpakete benutzbar wird. Das R-Commander-Paket ist ein klassisches Drag-and-Drop Menue auf dem R-Core (http://en.wikipedia. org/wiki/R_Commander) Das Statistiklabor ist ein spezielles Unterrichtsprogramm für die Statistik-Grundausbildung, das auf dem R-Core aufgebaut ist, vgl. Ghosh/Rendtel (2008) sowie Schlittgen (2004). Unausrottbar scheint der Versuch zu sein, das Tabellenkalkulationsprogramm EXCEL zu einem Statistik-Auswertungsprogramm umzufunktionieren. Immerhin wurde die Excel Methodenbank in dieser Hinsicht stark erweitert, vgl. hierzu auch Zwerenz (2008). Insbesondere kleinere Unternehmen, für die die Lizenz für ein klassisches Statistikpaket angeblich zu teuer ist, sind bereit, viel Geld auszugeben, wenn ihr Statistikproblem im Rahmen der Excel-Oberfläche lösbar wird. Auch dieses Problem wurde durch die R-Applikation R-in-Excel kostengünstig gelöst, vgl. Heiberger/Neuwirth (2009). Im Rahmen dieses Ansatzes verfügt Excel dann über ein eingebautes R-Fenster, innerhalb dessen die gesamte Funktionalität von R zur Verfügung steht. In diesem Sinne fehlt in SAS nur noch die Prozedur R, das Fenster zum R-Core. Angeblich ist das schon in Arbeit.

7. Prognosen über die Entwicklung der Statistik

Welche Trends können aus diesem kurzen Abriss der sozial- und wirtschaftswissenschaftlichen Statistik hergeleitet werden? Generell würde ich eine stärkere Datenorientierung und mehr Datennutzer verbunden mit einer verstärkten Nachfrage nach Beratung prognostizieren.

Weiterhin besteht ein Trend zu immer größeren Mikrodatensätzen. Dies bezieht sich sowohl auf die Fallzahl als auch auf den Umfang der Merkmale.[11] Mit gut 500.000 Beobachtungen im Mikrozensus stellt sich die Frage der Signifikanz von Ergebnissen ganz anders als beispielsweise bei einer Zeitreihe mit 100 Einzelbeobachtungen. Damit will ich hier nicht einer unbedarften Datenhuberei

11 Zwar ist der Anzahl der erfragten Merkmale eine natürliche Grenze durch die Belastbarkeit der Befragten gesetzt, aber durch die Kumulation von Merkmalen im Rahmen von Panel-Surveys ist der Anzahl der Merkmale pro Einheit kaum eine Grenze gesetzt.

das Wort reden. Es gilt immer das das GIGO-Verdikt: Garbage In, Garbage Out! Wenn ich in die Datenauswertung keine sinnvollen Fragen stecke, werde ich auch keine sinnvollen Antworten erhalten.

7.1 Future Data Access

Weiterhin würde ich mit einer noch stärkeren Verfügbarkeit von Daten rechnen. Dies betrifft zum einen die Neugründung von Forschungsdatenzentren, die zur Zeit wie Pilze aus dem Boden schießen (FDZ Nationales Bildungspanel, FDZ SOEP, drei FDZs bei GESIS). Zurzeit (Dez. 2010) sind 15 Forschungsdatenzentren sowie zwei Datenservicezentren auf der Internetseite des RatSWD gelistet. Auch die fachliche Basis des RatSWD verbreitert sich. So haben Epidemiologen, Demographen und Psychologen ihr Interesse und Engagement bekundet, im RatSWD mitzuarbeiten. Eine gewisse Bremse für den Datenzugang ist die Beschränkung auf Scientific-Use-Files. Dieses Weitergabeformat ist mit dem Begriff der „Faktischen Anonymität“ des Bundesstatistikgesetzes (§ 16, Absatz 6) verknüpft, der einen „unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskraft“ für die Deanonymisierung verlangt. Dies schließt insbesondere die Weitergabe von Regionalinformationen sowie detaillierte Branchenangaben bei Arbeitsmarktanalysen aus.

Die Lösung dieses Problems könnte ein neuer Datenzugang über das sogenannte Fernrechnen sein. Bei diesem Datenzugang werden die Daten zentral auf Rechnern des Datenbesitzers gehalten und der Wissenschaftler kann Auswertungen auf diesem Zentralrechner durchführen lassen. Von der Rechnerkapazität ist dies zumindest in Ländern wie Schweden und den Niederlanden kein Hindernis mehr. In Schweden hat man sich schon aus Effizienzgründen (ein einziger Datensatz, der zentral immer aktuell dokumentiert wird) für dieses Konzept entschieden (Projekt MONA (= Microdata Online Access)).

Zentral an diesem Ansatz ist die Trennung von Merkmalen, die für sich genommen eine Identifikation von Personen ermöglichen (z.B. Merkmale in Kombination mit Regionalmerkmalen), und dem Ergebnis, für das sich der Wissenschaftler interessiert und das er ggf. veröffentlichen möchte. In den allermeisten Fällen benötigt man die identifizierenden Merkmale nur für Zwischenschritte, z.B. die Regionaldaten zum Hinzufügen von Merkmalen des Wohnumfelds, und das Analyseergebnis ist in datenschutzrechtlicher Hinsicht völlig unbedenklich. Zur

Zeit wird diese Problematik von der AG „Future Data Access“ im RatSWD bearbeitet, vgl. http://www.ratswd.de/Future_Data_Access/index.php. Dies ist nur ein Beispiel für die Verzahnung von technologischer Entwicklung von Datenhaltung und –verarbeitung einerseits und Datenschutz andererseits.

Die letzte große Runde der Datenschutzgesetzgebung (1987) basierte auf der Auseinandersetzung mit dem Volkszählungsurteil von 1983. Diverse Datenschutzskandale im betrieblichen Bereich, aber auch die aktuellen Auseinandersetzungen um Google „Street View“ zeigen, dass der Datenschutzgesetzgebung eine gründiche Revision bevorsteht. Man kann nur hoffen, dass das Forschungsinteresse der Wissenschaft durch flexiblere Regelungen gestärkt wird. Sicher ist dies aber keinesfalls.

7.2 Vom Amt zur Agentur: Official Statistics

In der öffentlichen Wahrnehmung wird „die Statistik“ häufig mit der Amtlichen Statistik, jenem förderalen Konglomerat von Statistischem Bundesamt, 14 statistischen Landesämtern und noch mehr kommunalen Ämtern der Städtestatistik gleichgesetzt. Auf Bundesebene und meist auch auf Landesebene sind diese Ämter nachgeordnete Behörden der Innenministerien. Das hat Konsequenzen. Zum einen werden diese Ministerien mit Polizei und Terrorismusbekämpfung assoziiert. Wichtiger noch ist die Wahrnehmung der Statistik durch das Innenministerium selbst. Nach der Polizei taucht zu allererst der Sport (ebenfalls im Zuständigkeitsbereich des Innenministeriums) im Bewusstsein des Ministers auf. Der Sport ist in der öffentlichen Wahrnehmung ja auch viel wichtiger als die Statistik.[12] Wenn die Statistik im Ministerium überhaupt wahrgenommen wurde, dann unter dem Rubrum „Bürokratieabbau“. Falls im Rahmen von Terrorismuswarnungen zusätzliche Polizei- und Sicherheitskräfte aktiviert werden sollen, so hat dies in der Regel haushaltsneutral zu erfolgen. Also muss im Budget des Innenministers umgeschichtet werden. Und da ist der Personalbestand des Statistischen Bundesamts im Portfolio des Bundesinnenministers relativ groß und folglich muss er am stärksten mit Beiträgen zur Terrorismusabwehr rechnen.

12 Besonders fatal war die öffentliche Meinungsbildung im Zusammenhang mit der Revision des Hartz IV-Regelsatzes. Die Neuberechnung basierte auf den Daten des Statistischen Bundesamts, genauer gesagt auf den Daten der Einkommens- und Verbraucherstichprobe. Die „statistischen Tricks“, die dann im Ergebnis zu einer Erhöhung um € 5,-- führten, waren Vorgaben des Bundesministeriums für Arbeit und Soziales. So beziffert etwa Martens (2010) vom Paritätischen Wohlfahrtsverband den Wert dieser Tricks auf € 53,--. In der Öffentlichkeit äußerte sich das Statistische Bundesamt dazu mit keinem Wort. Offensichtlich fühlte man sich an die Kabinettsdisziplin des Ministeriums gebunden.

Es ist überraschend, wie freimütig Mitarbeiter der Statistikämter über Zukunftsvisionen für die Amtliche Statistik reden. Hier einige Leitlinien: Der föderalen Zersplitterung der Statistikämter und ihrer Abhängigkeit von dem Innenministerium wird das Modell einer nationalen, nichtabhängigen Statistik-Agentur entgegen gesetzt: mit eigenem Haushalt und einem wissenschaftlichen Beirat. Vergleichbar wäre diese Agentur in etwa mit der Bundesagentur für Arbeit. Das ist ein für Deutschland erprobtes Modell. Im internationalen Vergleich sind nationale Statistik-Agenturen sowieso die Regel.

Unterhalb dieser Stufe sind jedoch auch viele Detailänderungen wünschenswert. Die Etablierung des Forschungsdatenzentrums schließt auch den Begriff „Forschung“ ein. Bislang sucht man im Bundesstatistikgesetz vergeblich nach einer Verpflichtung zur Forschung. Nach bisherigem Verständnis sei man lediglich Datenproduzent. Dieses Selbstverständnis zeigt sich auch in der Organisationsstruktur des Amts. Es gibt im Statistischen Bundesamt lediglich zwei Gruppen, die sich mit Forschung beschäftigen, nämlich die Gruppe „Forschung und Entwicklung“ (Gruppe I b) und die Gruppe „Mathematisch Statistische Methoden“ (Gruppe II a). Dies sind zwei relativ kleine Methodengruppen, aufgesplittet auf zwei verschiedene Abteilungen. Daneben gibt es die großen „Fachabteilungen“: Gesamtrechnung/Arbeitsmarkt, Unternehmensregister/Industrie/Binnenhandel, Preise/Aussenhandel/Verkehr/Verdienste, Bevölkerung/Bildung/Staat, Dienstleistungen/Landwirtschaft/Umwelt und Gesundheit/Soziales/Private Haushalte. Allein dieser Stellenplan sagt etwas über das Verhältnis von Produktion und Forschung aus. Aber Stellenpläne und Abteilungsstrukturen kann man ja ändern!

Ein weiteres Essential von Forschung sind wissenschaftliche Aufsichtsgremien (Beiräte). Wer den „statistischen Beirat“ für eine solche Institution hält, der irrt. Es handelt sich vielmehr um ein Abstimmungsinstrument zur Koordination des Erhebungsprogramms der Amtlichen Statistik. Der Vorsitzende des Beirats ist der Präsident des Statistischen Bundesamts. Immerhin gibt es erste Ansätze zu wissenschaftlichen Beiräten. So wurde für den Zensus 2011 vom Innenminister ein wissenschaftlicher Beirat benannt, der eine solche Bezeichnung auch verdient.

Auch wenn in den statistischen Ämtern eher die Tendenz besteht, sich von seinesgleichen, sprich anderen nationalen statistischen Ämtern, evaluieren zu lassen, sollte man die Orientierung auf das akademische Umfeld nicht ignorieren. Man steht nämlich zusehends in Konkurrenz zu akademisch geleiteten Surveys:

Beispielsweise wurde die deutsche Teilstichprobe des Europäischen Haushaltspanels zunächst durch das Statistische Bundesamt erhoben. Nach drei Wellen wurde diese Erhebung zugunsten des akademischen Sozio-ökonomischen Panels (SOEP) eingestellt. Für die Zukunft ist es durchaus vorstellbar, dass die Amtliche Statistik Konkurrenz bekommt. Insbesondere bei den stetig wachsenden EU-weiten Surveys, die meist öffentlich ausgeschrieben werden, haben die Akademiker häufig die Nase vorn oder die Amtliche Statistik tritt erst gar nicht in den Wettbewerb ein.[13]

Das Verhältnis von Amtlicher Statistik und akademischem Umfeld hat sich durch die Einrichtung von Forschungsdatenzentren verbessert und die Beziehungen sind intensiviert worden. Dies war nicht immer so. Meist waren die Wissenschaftler frustiert über die Reaktionsgeschwindigkeit der Ämter gegenüber Daten- und Auswertungswünschen. Umgekehrt waren die Ämter als Abnehmer frischgebackener Akademiker mit dem Universitätsoutput unzufrieden: Eine zu starke Fixierung auf eine abstrakte Modellwelt zu Lasten von Defiziten beim Umgang mit empirischen Daten, vgl. Rendtel (2008).

Ist man Optimist, so kann sich in der Zukunft das Verhältnis von Amtlicher und Akademischer Statistik nachhaltig verbessern. Dies geschieht zum einen durch die Etablierung von spezifischen Masterabschlüssen, z.B. durch ein Masterprogramm in Survey Statistik (Bamberg/Berlin/Trier) oder in Survey Methodology (Duisburg).[14] Mit etwas Kooperation könnte man vielleicht auch ein spezielles Ausbildungsprogramm für „Official Statistics“, also die Amtliche Statistik etablieren, wie dies zurzeit bei Eurostat angedacht wird.

Etwas mehr Zeit wird wahrscheinlich die Bildung eines zentralen Methodenzentrums der Amtlichen Statistik in Deutschland dauern. Ein solches Zentrum könnte analog zum IAB (Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit) die Forschungstätigkeit der Amtlichen Statistik bündeln und intensivieren. Selbstredend wären die Leitungsstellen mit einer S-Professur an einer Universität gekoppelt, wie dies z.B. bei den Wirtschaftsforschungsinstituten üblich geworden ist. Um die akademische Lehre und Forschung enger mit der Amtlichen Statistik zu verzahnen, würden mehrere Stiftungsprofessuren einge-

13 Beispiele sind neben dem SOEP (DIW Berlin) das Nationale Bildungspanel NEPS (Universität Bamberg), der europaweite Alterssurvey SHARE (Universität Mannheim), das Partnerschafts- und Familienpanel PAIRFAM (Universität Mannheim) sowie der europaweite General Survey Social Sciences GSS (Gesis Mannheim).

14 Beide Masterprogramme sind im WS 2010/11 gestartet.

richtet. Vorbild wäre dabei die Stiftung Geld und Währung, die mit Mitteln der Bundesbank und des Finanzministeriums an der Goethe-Universität Frankfurt/Main drei Forschungsprofessuren[15] finanziert.

Aufgrund dieser Maßnahmen könnte sich das Verhältnis von Amtlicher Statistik und akademischer Statistik zum gegenseitigen Vorteil fruchtbar entwickeln.

7.3 Die Zukunft des Fachs Statistik in Zeiten von Bachelor und Master

Die Umstellung auf die Bachelor- und Masterstudiengänge hat nicht nur die Curricula innerhalb des Fachs Statistik durcheinander gewirbelt. Als Nebenfach mit geringer Hausmacht hatte die Statistik im Kampf um Repräsentation in den verkürzten Studiengängen ohnehin schlechte Karten. Hinzu kommt ein langfristiger Trend: In den 1980er-Jahren war der Anteil unter den nicht-statistischen Hochschullehrern mit eigenen Erfahrungen in der Datenanalyse und detaillierteren Statistikkenntnissen gering. Da benötigte man für die Ausbildung schon die gelernten Statistiker. Heute sieht das anders aus: Immer mehr Hochschullehrer der Wirtschaftswissenschaften haben selber empirische Datenanalyse betrieben und trauen sich zu, ihren Studierenden das benötigte statistische Wissen selbst beizubringen.

Schließlich ist zu bedenken, dass die Masterphase mit vier Semestern, wovon ein Semester für die Masterarbeit benötigt wird, nicht eben lang ist, um die gegebenenfalls verkürzte Statistikausbildung während der Bachelorphase wettzumachen und durch fortgeschrittene Kurse zu ergänzen. Eine Lösung könnte in einer Spezialisierung der Bachelor- und Masterstudiengänge liegen. Es findet sich schon jetzt eine bemerkenswerte Streuung der Studienpunkte im quantitativen Pflichtbereich von wirtschaftswissenschaftlichen Bachelorstudiengängen, vgl. Rendtel (2009). Um die Statistik als akademisches Fach zu erhalten, bietet sich die Einrichtung von spezialisierten Masterprogrammen an. Dies ist auch in einigen Fällen geschehen. Neben Dortmund und München, wo schon Diplomstudiengänge in Statistik existierten, wurde 2008 ein gemeinsames Masterprogramm Statistik an den drei Berliner Universitäten zusammen mit der medizinischen Hochschule in Berlin, der Charité, gestartet. Hinzu gekommen sind Masterprogramme in Statistik in Bielefeld und in Magdeburg, ein Survey-Master in Bamberg und

15 Die Stiftung hatte die Professuren für die Bereiche Jura, BWL und VWL für jeweils 10 Jahre ausgeschrieben. Die Goethe-Universität erhielt den Zuschlag.

Trier in Kooperation mit dem Berliner Statistik-Master sowie ein Master in Bio-Statistik in Bremen. Ich prognostiziere die Einrichtung weiterer Statistikstudiengänge auf Masterniveau. Und vielleicht realisiert sich auch noch der europäische Gedanke des Bologna-Prozesses: Die Etablierung von Ausbildungsprogrammen auf europäischer Ebene. Z.B. denkt man bei Eurostat über einen Master in Official Statistics nach.

7.4 Die Persistenz von Statistikpaketen

Die Nutzung des vertrauten Statistikpakets hat eine hohe Persistenz: Man wechselt ungerne eine viel benutzte Programmierumgebung. Nach einer gewissen Prägephase wird man nur selten beispielsweise von SAS zu SPSS wechseln. Weiterhin gibt es gewissen Wissenschaftsbereiche, wo bestimmte Programmpakete eine fast absolute Vorherrschaft haben: SPSS bei Psychologen und Soziologen, SAS bei Klinischen Studien, EVIEWS bei Ökonometrikern und STATA bei Panelanalysten. Wechseln tut nur der, der wechseln muss; z.B. weil das präferierte Programmpaket an der jeweiligen Universität oder Firma aus Kostengrünen nicht verfügbar ist. Es kommt gar nicht so selten vor, dass man das Statement „Paket XYZ können wir uns nicht leisten“ hört. Und dieses Argument ist seit der Existenz von R, das keine Lizenzgebühren verlangt, sehr schlagkräftig geworden. Damit wird R sich bei vielen Universitäten und kleineren Unternehmen über kurz oder lang durchsetzen. Bei den Großen sieht es aber anders aus. Diese haben ihren Betrieb auf ein bestimmtes Datenformat und eine Standardsoftware normiert und bei ihnen fallen die Lizenzgebühren gegenüber den Kosten einer Systemumstellung kaum ins Gewicht. Sie werden also auf jeden Fall bei ihrem Hausprogramm bleiben. Man denke etwa an diverse Betriebe, die trotz bitterer Klagen bei „ihrer SAP-Lösung“ bleiben, selbst wenn diese teuer und ineffizient ist. Die Statistik-Dinosaurier werden also nicht aussterben und diese werden ihrerseits versuchen, von R über geeignete Schnittstellen zu profitieren.

7.5 Die Statistikfelder der Zukunft

Die Überschrift „Die Zukunft der Statistik“ könnte die Erwartung wecken, ich unternähme den Versuch, die Statistikfelder der Zukunft zu skizzieren, frei nach dem Motto: „X wird In sein!“, „Y wird Out sein!“ Dies ist aber nur sehr

eingeschränkt möglich, will man sich nicht auf das dünne Eis der Spekulation begeben.

Immerhin kann man beobachten, dass die Finanzierung von Forschungsfragen einen nachhaltigen Einfluss auf das hat, was beforscht wird. Die Einflussnahme geschieht immer häufiger über privat finanzierte (Stiftungs)professuren an öffentlichen Universitäten. Da finanziert der Investmentverband für fünf Jahre eine Professur für „Investment und Altersvorsorge“ oder es wird ein Mitarbeiter für die „T-Mobile-Professur für M-Commerce“ gesucht. Das sind authentische Fälle!

In dieser Hinsicht dürfte die Zukunft von „Financial Statistics“ gesichert sein, solange sich die Banken von der Statistik eine Hilfe bei der Lösung ihrer Probleme erwarten. Bei anderen Statistikfeldern ist das weniger gut vorhersehbar. Der Einsatz von Pattern Recognition bei der Entschlüsselung von Genomen war ebenso wenig vorhersagbar wie die sozial-statistische Evaluation der Hartz-Gesetzgebung. Generell nimmt aber der Trend zur Datenanalyse in immer weiteren Anwendungsfeldern zu, so dass die Aussage, dass die Statistik eine Zukunft hat, als absolut sicher gelten kann!

8. Zusammenfassung

Aus der Retrospektive der letzten 30 Jahre war die Entwicklung der Statistik ausgesprochen dynamisch. Alle Parameter der Statistik, als der Wissenschaft von Daten und Denken, haben sich entscheidend verändert, meistens zum Vorteil. Allein die Fächerkonkurrenz an den Universitäten wird mit großer Härte ausgetragen. Die notorische Unterfinanzierung der deutschen Universitäten macht auch der Statistik zu schaffen, aber diese Mangelsituation gilt für die anderen Fächer genauso. Mit dem Rat für Sozial- und Wirtschaftsdaten (RatSWD) ist ein neues Element entstanden, das auf absehbare Zeit die Beziehungen zwischen Datenproduzenten und Datennutzern intensivieren wird. Dies wird insbesondere die Beziehungen zwischen der Amtlichen und der Akademischen Statistik zum gegenseitigen Nutzen intensivieren.

Literatur:

Cheng, B.; Titterington, D.M. (1994): Neural Networks: A Review from a Statistical Perspective. Statistical Science, 9, S. 2-30.

Davison, A.C.; Hinkley, D.V. (1997): Bootstrap Methods and their Applications, Cambridge University Press, Cambridge.

Efron, b.; Tibshirani, R. (1993): An Introduction to the Bootstrap, Chapman and Hall, New York.

Fisher, R.A. (1925): „Theory of Statistical Estimation“, Proceedings of the Cambridge Philological Society, 22, S. 700-725.

Gelman, A.; Meng, X. (2004): Applied Bayesian Modeling and Casual Inference from Incomplete-Data Perspectives. Wiley, New York.

Ghosh, A.; Rendtel, U. (2008): Unterrichten und Prüfen mit dem Statistiklabor: Ein Erfahrungsbericht. ASTA Wirtschafts- und Sozialstatistisches Archiv, 2, S. 145 – 164.

Hastie, T.; Tibshirani, R. (2001): The Elements of Statistical Learning. Springer, Heidelberg.

Heiberger, R.; Neuwirth, E. (2009): R Through EXCEL. Springer, Heidelberg

Lechner, M.; Pfeiffer, F. (eds.) (2007): Econometric Evaluation of Labour Market Policies. Physika Verlag, Heidelberg.

Martens, R. (2010): Die Hartz IV-Abrechnung: Blätter für deutsche und internationale Politik, Heft 11/2010, S. 5-8.

Matignon, R. (2007): Data Mining Using SAS Enterprise Miner. Wiley, New York.

Raghunathan, T.E., Solenberger, P.; v. Hoewyk, J. (2002): IVEware: Imputation and Variance Estimation Software, User Guide. Survey Research Center, University of Michigan.

Rendtel, U. (2008): Statistikausbildung und Amtliche Statistik. Kritik und Perspektiven. ASTA Wirtschafts- und Sozialstatistisches Archiv, 2, S. 5 -20.

Rendtel, U. (2009): Teaching and Statistical Training, Working Paper No. 81. German Council for Social and Economic Data, RatSWD, Berlin, http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_81.pdf .

Rubin, D. (1996): Multiple Imputation after 18+ Years. JASA, 91, S. 473 – 489.

Schlittgen, R. (2004): Das Statistiklabor. Einführung und Benutzerhandbuch. Springer, Heidelberg.

Schmid, F.; Trede, M. (2005): Finanzmarktstatistik, Springer.

v. Buuren, S.; Groothuis-Oudshoorn, K. (2010): MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, forthcoming.

v.d. Lippe, P.; Kladobra, A. (2008): Der unaufhaltsame Niedergang der Fächer Statistik und Ökonometrie in den Wirtschaftsiwssenschaften. ASTA Wirtschafts- und Sozialstatistisches Archiv, 2, S. 21-40.

Wagner, G.; Büning, H. (2008): Statistik als Instrument zum Hinausprüfen von Studierenden?, ASTA Wirtschafts- und Sozialstatistisches Archiv, 2. S. 51 -54.

Zwerenz, K.-H. (2008): E-Learning-Szenarien für die Statistikausbildung: Virtuelle hochschule und Live-E-Learning. ASTA Wirtschafts- und Sozialstatistisches Archiv, 2, S. 165–174.

Zwick, M. (2008): Campus-Files – Kostenfreie Public Use Files für die Lehre, ASTA Wirtschafts- und Sozialstatistisches Archiv, 2. S. 175-188.